**The promise of satellite monitoring for preserving biodiversity** p. 926

**Antiaging diets dissected** p. 953

**How a viral RNA hijacks host machinery** p. 955

# Science

$15
19 NOVEMBER 2021
science.org

AAAS

## JUPITER'S STORMS

Juno peers into the atmosphere

pp. 964 & 968

VISIT…

LANZAROTE
Caliente.COM

**Gordon Research Conferences**

# New Conference Application Process

Join the over 400+ GRC communities and experience a high-quality scientific conference experience while also advancing the frontiers in your field!

## *New Conferences in 2022*

- Additive Manufacturing of Soft Materials
- Biology of Winter
- Calcium Signaling
- Cell-Cell Fusion
- Computational Materials Science and Engineering
- Electrochemical Interfaces in Energy Conversion and Storage
- Fluids in Disease Transmission and Contamination
- Frontal Cortex
- Genetic Biocontrol
- Graft Preservation in Heart Transplantation
- Hypothalamus
- Immunoengineering
- Immunometabolism in Health and Disease
- In Vivo Ultrasound Imaging
- Lipidomics
- Multiscale Mechanochemistry and Mechanobiology
- Neural Mechanisms of Acoustic Communication
- Radionuclide Theranostics for the Management of Cancer
- Single-Cell Cancer Biology
- Single-Cell Genomics
- Systems Aging
- Unconventional Semiconductors and Their Applications

**Applications for 2022 conferences are being accepted at www.grc.org**

Is your scientific community interested in gathering as a small group to discuss cutting-edge, frontier science while also experiencing high-quality, intimate and informal interactions and networking opportunities with peers from around the world? Consider submitting a proposal to initiate a brand-new GRC!

GRC is always interested in receiving new proposals for cutting-edge topics not already covered by the GRC portfolio.

The GRC Board is particularly interested in proposals from engineers and computer scientists!

Proposals are due on an annual basis on June 1 and interested individuals should reach out to Beth Mancuso (**bmancuso@grc.org**) for more information!

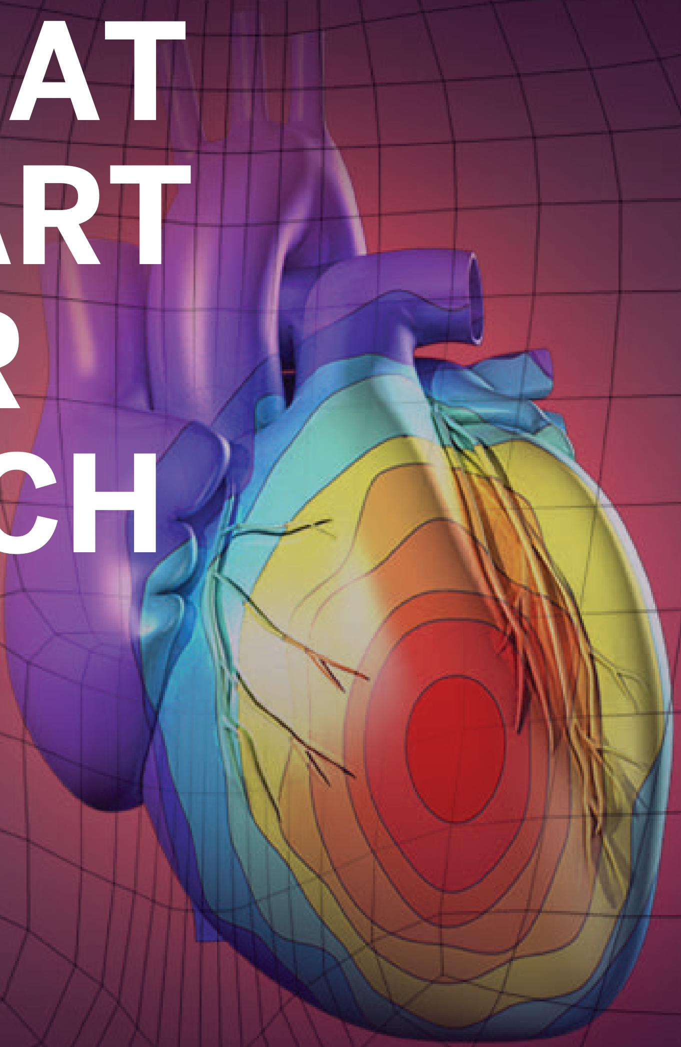
ScienceTranslationalMedicine.org

PUT HUMAN HEALTH AT THE HEART OF YOUR RESEARCH

Submit your research:
**cts.ScienceMag.org**

Science
Translational
Medicine
AAAS

Twitter: @ScienceTM
Facebook: @ScienceTranslationalMedicine

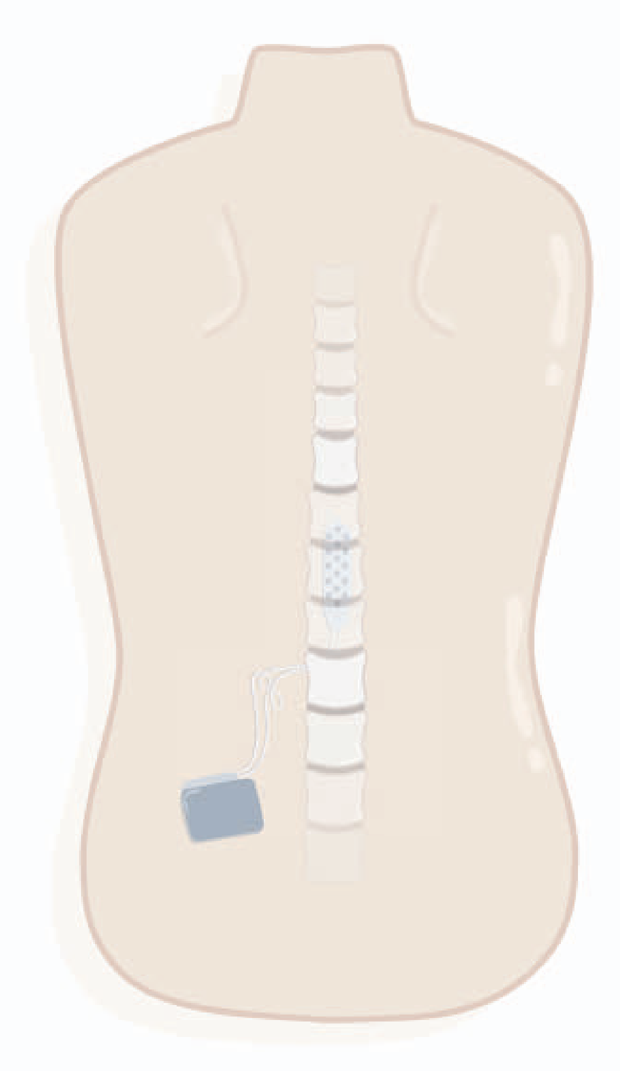


# Neuromodulation improves quality of life after spinal-cord injuries

*Treating pain with pulses of electricity started centuries ago, but modern technology optimizes the methods for using such treatments, especially for patients with an injured spinal cord. Beijing PINS Medical Co., Ltd.'s neuromodulating tools provide physicians with exquisite control of these treatments.*

Around the world, as many as half a million people injure their spinal cord every year. According to the World Health Organization, people with a spinal-cord injury (SCI) are as much as five times more likely to die prematurely than people without one. Advanced treatments, including neuromodulation, can dramatically improve the lives of these patients.

"Spinal-cord injury patients partly or completely lose their sensory, motor, and some other functions below the injury level," explains Yanan Sui, associate professor in the National Engineering Laboratory for Neuromodulation at Tsinghua University in Beijing, China. Surgically decompressing the injury within 24 hours "can restore the biomechanical structure and stability of the spine," Sui notes, and "neuroprotective drugs can reduce the primary and secondary injury to preserve the residual function of the spinal cord."



An epidural electrical stimulation system used in Yanan Sui's research (*1*) to treat spinal cord injury.

Following treatments after the injury, a patient will need rehabilitation. Neuromodulation offers significant relief to some SCI patients—and hope for the future.

## Pulses to reduce pain

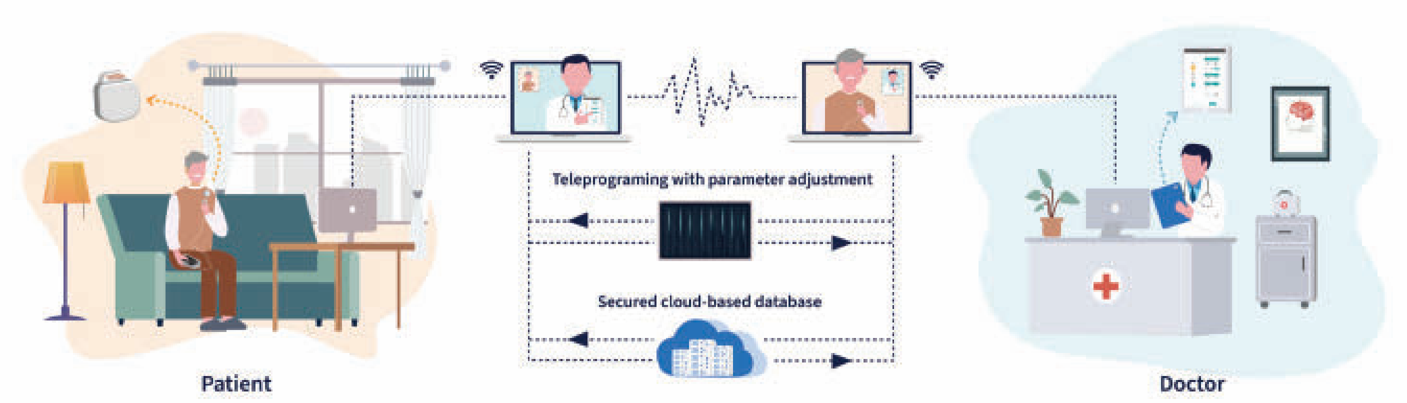
More than half of SCI patients experience ongoing pain. For centuries, physicians have treated pain with electrical stimulation. In fact, as early as the first century AD, ancient Roman physicians prescribed sitting in a pool of water that contained torpedo fish—also known as electric rays—which provided pulses of electricity to ease a patient's pain. Pulses of electricity can also reduce the pain of SCI patients, but modern medicine eliminates the need for fish. Instead, today's physicians can use epidural electrical stimulation (EES).

To understand the potential impact of neuromodulation after an SCI, consider the consequences of such an injury. "After an SCI, the spinal cord lacks the descending information from the motor cortex, and the interruption of the motor circuit results in the loss of cortical motor control of the lower limbs," Sui explains. "Although the spinal cord is intact below the level of injury, its own neural circuit is inactivated." EES can provide that activation.

According to Yang Lu, a neurosurgeon at Beijing Tsinghua Changgung Hospital (an affiliated hospital of Tsinghua University) who studies spinal surgery and spinal-cord neuromodulation, "Most cases in our center who use the spinal-cord neuromodulation therapies are intractable pain patients." This therapy is usually a two-step procedure. "Percutaneous or paddle leads are chosen based on the individual conditions," Lu says. "The first step is a trial phase in which the patient is stimulated with an external pulse generator to test the clinical outcome." If the patient experiences 50% or more improvement, a pulse generator is implanted to continue the neuromodulatory treatment.

Sui agrees with the value of EES. "Epidural electrical stimulation of the spinal cord—a traditional neuromodulation therapy—is one of the most promising strategies to treat SCI," he says. "Its striking clinical effect has been demonstrated by several famous medical centers." He

IMAGE: PROVIDED BY PINS


Teleprograming with parameter adjustment
Secured cloud-based database
Patient
Doctor

The process for remote adjustment of a neuromodulation device.

adds, "Epidural stimulation in the lumbar enlargement area can induce rhythmic lower-limb movements and enable voluntary movements in SCI patients."

## Optimizing the technology

Relaxing in a pool of electric fish might provide some pain relief, but that treatment is about as far from controlled as it can be. Today's options–including technology from PINS Medical–give physicians precise control of neuromodulatory treatments. Sui calls PINS Medical "an innovative enterprise in the field of neuromodulation."

The features of PINS devices that really impress Sui include:

- neuromodulation with variable-frequency stimulation
- high-density electrodes and flexible control patterns for spinal-cord stimulation that provide high-accuracy stimulation
- advanced artificial intelligence algorithms that can find the optimal biomimetic neuromodulation solution.

According to Lu, PINS offers "the first remote-programming platform that allows physicians to remotely adjust the configurations and parameters of the stimulation." The feature adds important options for long-distance therapy. "China has a vast territory and a great population," Lu says. "However, the unbalanced medical resources create a hard challenge for the patients from relatively underdeveloped areas." With the PINS technology, though, a patient can "undergo surgery in a big center and then return home and receive the postoperative programming," he says.

Remote programming proved particularly useful during COVID-19. As Lu explains, "It was hard for patients in other places to go to the hospitals in Beijing or Shanghai because of the restrictions of social contact." Lu and his colleagues even published a paper on using spinal-cord stimulation with PINS technology to reduce the pain of SCI patients during the pandemic (*2*).

IMAGE: PROVIDED BY PINS

## Treating pain and paralysis

Neuromodulation improves the lives of many patients after an SCI. Lu describes a 56-year-old woman who experienced pain in both legs even 3 years after injuring her spinal cord. Despite surgery designed to address the pain, she still described a continuous burning sensation–giving it a 9 out of 10 for pain–and no medication helped. So, Lu and his colleagues implanted two percutaneous electrodes between her 9th and 10th thoracic vertebrae. "In the trial phase, careful searching for the sweetest spot of the stimulation could give over 80% pain reduction," Lu says. "Then, the patient accepted the implantation of the pulse generator."

After going through her injury and then 3 years of pain, this woman felt relief and gratitude. "She really thanked us, because what we did has changed her life," Lu says. "She said she could get a good sleep after 3 years of sleep deprivation."

In addition to reducing pain, neuromodulation might do even more for some SCI patients. That is part of Sui's vision for this technology. He says, "I worked with clinical experts to build a new type of treatment system." For example, he developed new theories in machine learning and online methods that use feedback from patients to reinforce learning. He adds, "Together with my colleagues, I applied the advanced machine-learning methods to clinical neuromodulation, treating patients paralyzed by spinal-cord injury through epidural electrical stimulation." These methods show great promise. So far, Sui says, this has "been successfully used in clinical treatment in hospitals, to help lower-limb paraplegic patients stand up again and to restore the hand-grip function of tetraplegic patients."

Even more opportunities lie ahead for applying neuromodulation to improve the lives of patients after SCIs. As Sui says, "At Tsinghua, my research group continues to collaborate with hospitals to perform translational medicine research on neuromodulation and to serve patients with neurological diseases."

## References

1. www.yanansui.com.
2. Y. Lu *et al.*, *Front. Neurosci.* **14**, 594696 (2020).

**Sponsored by**

PINS 品驰

# CONTENTS

19 NOVEMBER 2021 • VOLUME 374 • ISSUE 6570

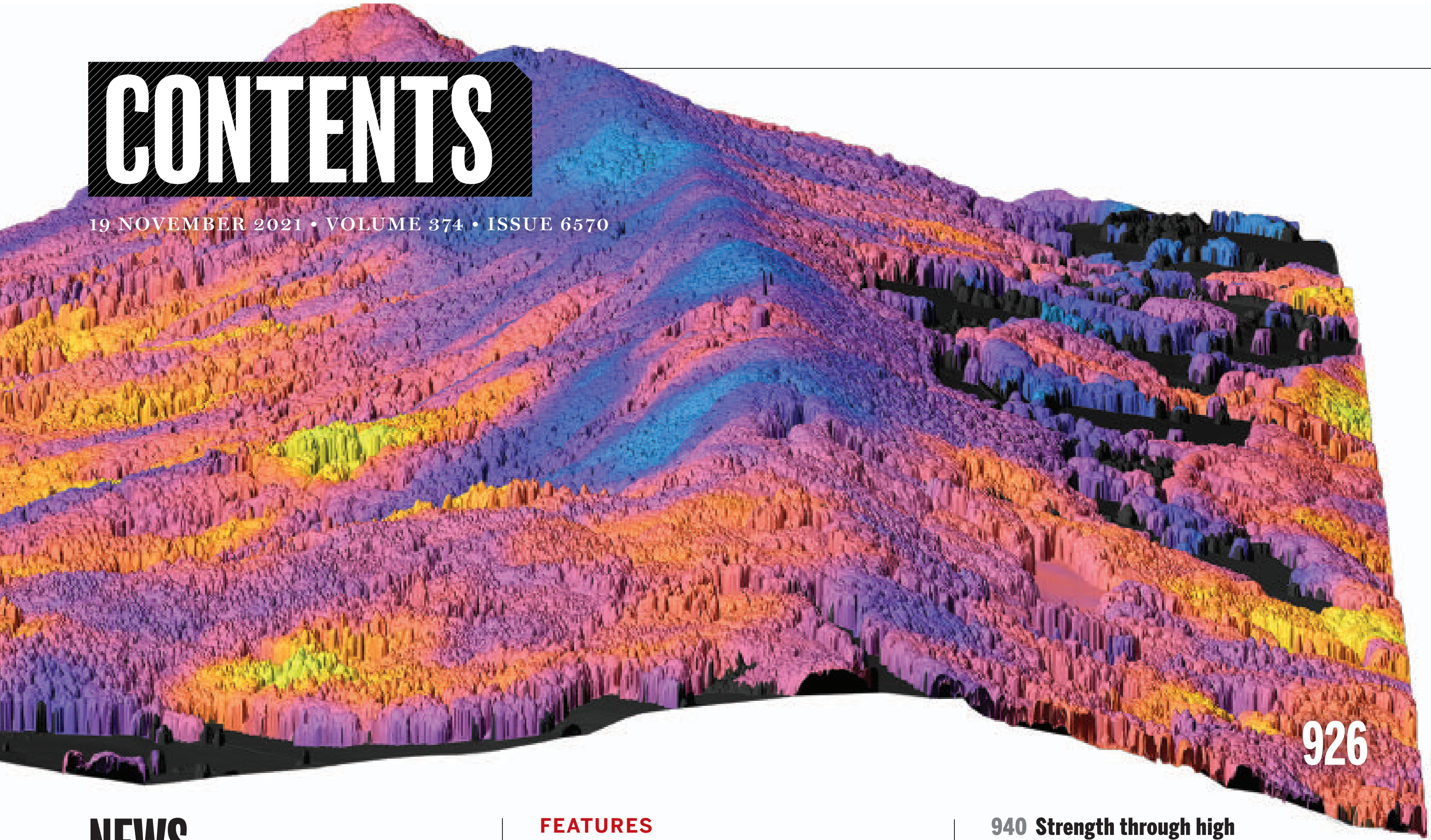


IMAGE: F. MORSDORF, *REMOTE SENSING OF PLANT BIODIVERSITY*, DOI.ORG/10.1007/978-3-030-33157-3_4 (2020)

# NEWS

## IN BRIEF

**918** News at a glance

## IN DEPTH

**920 Glasgow pact leaves 1.5°C goal on life support**
Goal of preventing dangerous warming slipping away, but nations boost emissions cuts *By C. O'Grady*

**921 High-flying wildfire smoke poses potential threat to ozone layer**
Record Arctic ozone loss linked to Siberian wildfires *By P. Voosen*
PODCAST

**922 Mediterranean mussel faces 'imminent extinction'**
After devastating epidemic, Europe launches new efforts to save one of the world's largest bivalves *By E. Stokstad*

**924 'Zero COVID' is getting harder—but China is sticking with it**
Government aims to increase immunity further before relaxing controls and trying to live with the virus *By D. Normile*

**925 To fight Lyme disease, vaccine targets ticks**
Ticks fed poorly, transmitted pathogen less often to guinea pigs given mRNA vaccine *By M. Wadman*

## FEATURES

**926 Getting the big picture of biodiversity**
Satellites and other remote sensing tools offer new ways to study ecosystems—and maybe even save them *By E. Pennisi*

# INSIGHTS

## POLICY FORUM

**932 Create a COVID-19 commission**
We need a definitive public reference for the history of events
*By C. F. Chyba* et al.

## PERSPECTIVES

**936 No vacancy in the Fermi sea**
Optical transparency is increased by the Pauli principle in ultracold atom gases
*By B. DeMarco and J. H. Thywissen*
REPORTS pp. 972, 976, & 979

**937 The head and the heart of fear**
The insular cortex calibrates brain, body, and behavioral responses to danger
*By J. P. Christianson*
REPORT p. 1010

**938 A molecular double-slit experiment**
Quantum interaction similar to light interference is observed in molecular collisions *By X. Wang and X. Yang*
RESEARCH ARTICLE p. 960

**940 Strength through high slip-plane density**
Subjecting a multicomponent alloy to cyclic torsion can create a strong, ductile material *By J.-W. Yeh*
REPORT p. 984

**941 Untangling the evolution of dengue viruses**
The push and pull of dengue virus serotype evolution influences epidemic potential
*By P. Rohani and J. M. Drake*
REPORT p. 999

**942 Digging into X chromosome inactivation**
Characterizing X chromosome inactivation in nonhuman primates reveals some surprises *By E. Heard and C. Rougeulle*
RESEARCH ARTICLE p. 954

## BOOKS *ET AL.*

**944 Shrinking snowpacks, anxious snow lovers**
A cold-weather enthusiast confronts connections between the cryosphere and human culture *By M. Serreze*

**945 Incomplete Indigenous landscapes**
A new text confronts how bias tainted efforts to preserve native cultures of the Americas *By P. Steeves*

## LETTERS

**947 Minimizing fear of wildlife in urban areas**
*By S. A. Lambertucci* et al.

**948 Surviving as a young scientist in Brazil**
*By C. H. L. Silva Junior* et al.

**948 Scientific denialism threatens Brazil**
*By L. M. Diele-Viegas* et al.

**949 Technical Comment abstracts**

# RESEARCH

## IN BRIEF

**950** From *Science* and other journals

## REVIEW

**953 Aging**
Antiaging diets: Separating fact from fiction *M. B. Lee* et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABE7365

## RESEARCH ARTICLES

**954 Developmental biology**
The X chromosome dosage compensation program during the development of cynomolgus monkeys *I. Okamoto* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD8887
PERSPECTIVE p. 942

**955 Virology**
A viral RNA hijacks host machinery using dynamic conformational changes of a tRNA-like structure *S. L. Bonilla* et al.

**960 Quantum mechanics**
Quantum mechanical double slit for molecular scattering *H. Zhou* et al.
PERSPECTIVE p. 938

## REPORTS

### Jupiter

**964** The depth of Jupiter's Great Red Spot constrained by Juno gravity overflights *M. Parisi* et al.

**968** Microwave observations reveal the deep extent and structure of Jupiter's atmospheric vortices *S. J. Bolton* et al.

### Quantum gases

**972** Observation of Pauli blocking in light scattering from quantum degenerate fermions *A. B. Deb and N. Kjærgaard*

**976** Pauli blocking of light scattering in degenerate fermions *Y. Margalit* et al.

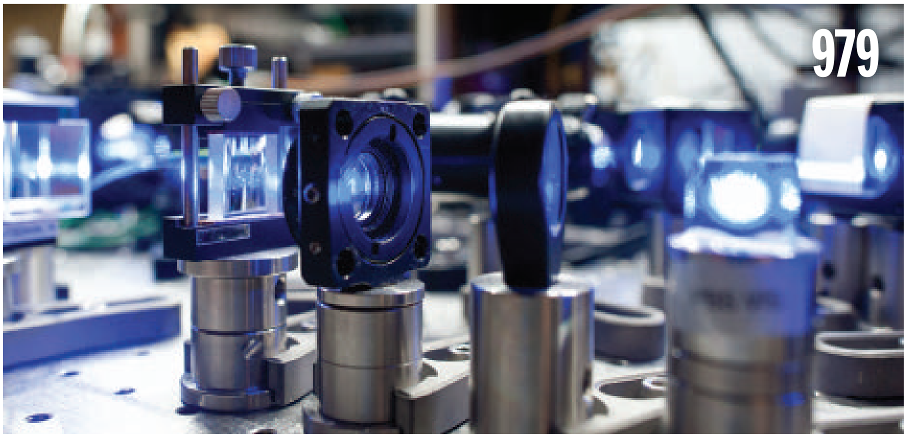
**979** Pauli blocking of atom-light scattering *C. Sanner* et al.
PERSPECTIVE p. 936

**984 Metallurgy**
Gradient cell–structured high-entropy alloy with exceptional strength and ductility *Q. Pan* et al.
PERSPECTIVE p. 940

**989 Coronavirus**
COVID-19 transmission dynamics underlying epidemic waves in Kenya *S. P. C. Brand* et al.

**995 Coronavirus**
Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India *M. S. Dhar* et al.

**999 Dengue evolution**
Antigenic evolution of dengue viruses over 20 years *L. C. Katzelnick* et al.
PERSPECTIVE p. 941



**Blue laser light is used to show quantum statistics effects on light scattering in an ultracold Fermi gas.**

**1005 Natural products**
Biosynthesis of fluopsin C, a copper-containing antibiotic from *Pseudomonas aeruginosa* *J. B. Patteson* et al.

**1010 Neuroscience**
Fear balance is maintained by bodily feedback to the insular cortex in mice *A. S. Klein* et al.
PERSPECTIVE p. 937

## DEPARTMENTS

**913 Editorial**
COVID-19 vaccines for children
*By Jeffrey S. Gerber and Paul A. Offit*

**1018 Working Life**
The hard way
*By Rassim Khelifa*

## ON THE COVER

Optical image of Jupiter's atmosphere, taken by the Juno spacecraft. The cloud top colors vary because of differing compositions, probably mixtures of water, ammonia, and ammonium hydrosulfide. The large elongated storm in the lower left is known as a barge; a more circular white storm is visible at the top right. Microwave and gravitational field data from Juno constrain the depths of the gas giant's storms. See pages 964 and 968. *Credit: NASA/JPL-Caltech/SwRI/MSSS/Kevin M. Gill*

Science
STORMS

*Science* Staff .......................................... 912
*Science* Careers .................................... 1016

CREDITS: (PHOTO) CHRISTIAN SANNER, YE LABS/JILA; (ILLUSTRATION) ROBERT NEUBECKER

**SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005.** Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2021 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
**Change of address:** Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

# Science AAAS

**Editor-in-Chief** Holden Thorp, hthorp@aaas.org

**Executive Editor** Monica M. Bradford

**Editors, Research** Valda Vinson, Jake S. Yeston **Editor, Insights** Lisa D. Chong

**DEPUTY EDITORS** Julia Fahrenkamp-Uppenbrink (UK), Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden (UK) **SR. EDITORS** Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, Di Jiang, Priscilla N. Kelly, Marc S. Lavine (Canada), Yevgeniya Nusinovich, Ian S. Osborne (UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Yuen Yiu, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Bianca Lopez, Seth Thomas Scanlon (UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Julia Haber-Katris, Nida Masiulis, Abigail Shashikanth, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Hilary Stewart (UK), Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Alexander Kief, Ronmel Navas, Isabel Schnaidt, Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock (UK)

**News Editor** Tim Appenzeller

**NEWS MANAGING EDITOR** John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Kelly Servick, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Rodrigo Pérez Ortega (Mexico City), Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERNS** Menna Ibrahim, Sofia Moutinho, Anil Oza, Alex Viveros **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Andrew Curry (Berlin), Ann Gibbons, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Tokyo), Elisabeth Pain (Careers), Charles Piller, Michael Price, Tania Rabesandratana (Barcelona), Joshua Sokol, Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Rachel Bernstein (Editor), Katie Langin (Associate Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Morgan Everett, Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

**Creative Director** Beth Rakouskas

**DESIGN MANAGING EDITOR** Marcy Atarod **INTERIM GRAPHICS MANAGING EDITOR** Chris Bickel **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB STRATEGY MANAGER** Kara Estelle-Powers **MULTIMEDIA MANAGING PRODUCER** Joel Goldberg **DESIGN EDITOR** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS EDITOR** Nirja Desai **INTERACTIVE GRAPHICS EDITOR** Kelly Franklin **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR SCIENTIFIC ILLUSTRATOR** Valerie Altounian **SCIENTIFIC ILLUSTRATOR** Ashley Mastin **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **SOCIAL MEDIA PRODUCER** Sabrina Jenkins **WEB DESIGNER** Jennie Pajerowski **SENIOR PODCAST PRODUCER** Sarah Crespi **VIDEO PRODUCER** Meagan Cantwell

**Chief Executive Officer and Executive Publisher** Sudip Parikh

**Publisher, Science Family of Journals** Bill Moran

**DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS** Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **MANAGER, BUSINESS ANALYSIS** Cory Lipman **BUSINESS ANALYST** Kurt Ennis **FINANCIAL ANALYST** Isacco Fusi **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT AND PUBLISHING SYSTEMS SPECIALIST** Jacob Hedrick **SENIOR CONTENT SPECIALISTS** Steve Forrester, Lori Murphy **PRODUCTION SPECIALIST** Kristin Wowk **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

**ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT** Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Ashley Hylton, Mike Romano, Tori Velasquez, Jenna Voris, Justin Wood **SENIOR DESIGNER** Kim Huynh

**DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING** Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst

**DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT** Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **PUBLISHING TECHNOLOGY MANAGER** Michael Di Natale **SR. PRODUCT ASSOCIATE** Robert Koepke **PRODUCT ASSOCIATE** Anne Mason **SPJ ASSOCIATE MANAGER** Samantha Bruno Fuller **SPJ ASSOCIATE** Casey Buchta

**MARKETING MANAGER** Kess Knight **BUSINESS DEVELOPMENT MANAGER** Rasmus Andersen **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan, Claudia Paulsen-Young **SENIOR MANAGER, INSTITUTIONAL LICENSING OPERATIONS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer

**DIRECTOR, GLOBAL SALES** Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor **US MID WEST, MID ATLANTIC AND SOUTH EAST SALES** Chris Hoag **US WEST COAST SALES** Lynne Stickrod **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Victoria Glasbey **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **MARKETING MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Yoshimi Toda (Tokyo), Miyuki Tani (Osaka)

**DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS** Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **LICENSING ASSOCIATE** Virginia Warren

---

**MAIN HEADQUARTERS**
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

**SCIENCE INTERNATIONAL**
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

**SCIENCE CHINA**
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

**SCIENCE JAPAN**
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

**EDITORIAL**
science_editors@aaas.org

**NEWS**
science_news@aaas.org

**INFORMATION FOR AUTHORS**
science.org/authors/science-information-authors

**REPRINTS AND PERMISSIONS**
science.org/help/reprints-and-permissions

**MEDIA CONTACTS**
scipak@aaas.org

**MULTIMEDIA CONTACTS**
SciencePodcast@aaas.org
ScienceVideo@aaas.org

**INSTITUTIONAL SALES AND SITE LICENSES**
science.org/librarian

**PRODUCT ADVERTISING & CUSTOM PUBLISHING**
advertising.science.org/products-services
science_advertising@aaas.org

**CLASSIFIED ADVERTISING**
advertising.science.org/science-careers
advertise@sciencecareers.org

**JOB POSTING CUSTOMER SERVICE**
employers.sciencecareers.org
support@sciencecareers.org

**MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS**
science.org/subscriptions

**MEMBER BENEFITS**
aaas.org/membership/benefits

**AAAS BOARD OF DIRECTORS**
**CHAIR** Claire M. Fraser
**PRESIDENT** Susan G. Amara
**PRESIDENT-ELECT** Gilda A. Barabino
**TREASURER** Carolyn N. Ainslie
**CHIEF EXECUTIVE OFFICER** Sudip Parikh
**BOARD** Cynthia M. Beall
Rosina M. Bierbaum
Ann Bostrom
Janine Austin Clayton
Laura H. Greene
Kaye Husbands Fealing
Maria M. Klawe
Robert B. Millard
William D. Provine

*Science* serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

**BOARD OF REVIEWING EDITORS** (Statistics board members indicated with **S**)

**Erin Adams,** *U. of Chicago*
**Takuzo Aida,** *U. of Tokyo*
**Leslie Aiello,** *Wenner-Gren Fdn.*
**Deji Akinwande,** *UT Austin*
**Judith Allen,** *U. of Manchester*
**Marcella Alsan,** *Harvard U.*
**Sebastian Amigorena,** *Inst. Curie*
**James Analytis,** *UC Berkeley*
**Trevor Archer,** *NIEHS, NIH*
**Paola Arlotta,** *Harvard U.*
**David Awschalom,** *U. of Chicago*
**Delia Baldassarri,** *NYU*
**Nenad Ban,** *ETH Zürich*
**Nandita Basu,** *U. of Waterloo*
**Franz Bauer,** *Pontificia U. Católica de Chile*
**Ray H. Baughman,** *UT Dallas*
**Carlo Beenakker,** *Leiden U.*
**Yasmine Belkaid,** *NIAID, NIH*
**Philip Benfey,** *Duke U.*
**Kiros T. Berhane,** *Columbia U.*
**Joseph J. Berry,** *NREL*
**Alessandra Biffi**, *Harvard Med.*
**Chris Bowler,** *École Normale Supérieure*
**Ian Boyd,** *U. of St. Andrews*
**Emily Brodsky,** *UC Santa Cruz*
**Ron Brookmeyer,** *UCLA* (**S**)
**Christian Büchel,** *UKE Hamburg*
**Dennis Burton,** *Scripps Res.*
**Carter Tribley Butts,** *UC Irvine*
**György Buzsáki**, *NYU School of Med.*
**Mariana Byndloss,** *Vanderbilt U. Med. Ctr.*
**Annmarie Carlton,** *UC Irvine*
**Simon Cauchemez,** *Inst. Pasteur*
**Ling-Ling Chen,** *SIBCB, CAS*
**M. Keith Chen,** *UCLA*
**Zhijian Chen**, *UT Southwestern Med. Ctr.*
**Ib Chorkendorff,** *Denmark TU*
**Amander Clark,** *UCLA*
**James J. Collins,** *MIT*
**Robert Cook-Deegan,** *Arizona State U.*
**Virginia Cornish** *Columbia U.*
**Carolyn Coyne,** *Duke U.*
**Roberta Croce,** *VU Amsterdam*
**Ismaila Dabo,** *Penn State U.*
**Jeff L. Dangl,** *UNC*
**Chiara Daraio,** *Caltech*
**Nicolas Dauphas,** *U. of Chicago*
**Christian Davenport,** *U. of Michigan*
**Frans de Waal,** *Emory U.*
**Claude Desplan,** *NYU*
**Sandra Díaz,** *U. Nacional de Córdoba*
**Ulrike Diebold**, *TU Wien*
**Stefanie Dimmeler,** *Goethe-U. Frankfurt*
**Hong Ding,** *Inst. of Physics, CAS*
**Dennis Discher,** *UPenn*
**Jennifer A. Doudna,** *UC Berkeley*
**Ruth Drdla-Schutting,** *Med. U. Vienna*
**Raissa M. D'Souza,** *UC Davis*
**Bruce Dunn,** *UCLA*
**William Dunphy,** *Caltech*
**Scott Edwards,** *Harvard U.*
**Todd Ehlers,** *U. of Tübingen*
**Andrea Encalada,** *U. San Francisco de Quito*
**Nader Engheta,** *UPenn*
**Karen Ersche,** *U. of Cambridge*
**Beate Escher,** *UFZ & U. of Tübingen*
**Barry Everitt,** *U. of Cambridge*
**Vanessa Ezenwa,** *U. of Georgia*
**Michael Feuer,** *GWU*
**Toren Finkel,** *U. of Pitt. Med. Ctr.*
**Gwenn Flowers,** *Simon Fraser U.*
**Peter Fratzl,** *Max Planck Inst. Potsdam*
**Elaine Fuchs,** *Rockefeller U.*
**Jay Gallagher,** *U. of Wisconsin*
**Daniel Geschwind,** *UCLA*
**Ramon Gonzalez,** *U. of South Florida*
**Sandra González-Bailón,** *UPenn*
**Nicolas Gruber,** *ETH Zürich*
**Hua Guo,** *U. of New Mexico*
**Taekjip Ha,** *Johns Hopkins U.*
**Sharon Hammes-Schiffer,** *Yale U.*
**Wolf-Dietrich Hardt,** *ETH Zürich*
**Louise Harra,** *U. Coll. London*
**Carl-Philipp Heisenberg,** *IST Austria*
**Ykä Helariutta,** *U. of Cambridge*
**Janet G. Hering,** *Eawag*
**Christoph Hess,** *U. of Basel & U. of Cambridge*
**Heather Hickman,** *NIAID, NIH*
**Hans Hilgenkamp,** *U. of Twente*
**Kai-Uwe Hinrichs,** *U. of Bremen*
**Deirdre Hollingsworth,** *U. of Oxford*
**Randall Hulet,** *Rice U.*
**Auke Ijspeert,** *EPFL*
**Akiko Iwasaki,** *Yale U.*
**Stephen Jackson,** *USGS & U. of Arizona*
**Erich Jarvis,** *Rockefeller U.*
**Peter Jonas,** *IST Austria*
**Matt Kaeberlein,** *U. of Wash.*
**William Kaelin Jr.,** *Dana-Farber Cancer Inst.*
**Daniel Kammen,** *UC Berkeley*
**Kisuk Kang,** *Seoul Nat. U.*
**Sabine Kastner,** *Princeton U.*
**V. Narry Kim,** *Seoul Nat. U.*
**Robert Kingston,** *Harvard Med.*
**Nancy Knowlton,** *Smithsonian Institution*
**Etienne Koechlin,** *École Normale Supérieure*
**Alex L. Kolodkin,** *Johns Hopkins U.*
**Julija Krupic,** *U. of Cambridge*
**Paul Kubes,** *U. of Calgary*
**Gabriel Lander,** *Scripps Res.* (**S**)
**Mitchell A. Lazar,** *UPenn*
**Wendell Lim,** *UCSF*
**Luis Liz-Marzán,** *CIC biomaGUNE*
**Omar Lizardo,** *UCLA*
**Jonathan Losos,** *Wash. U. in St. Louis*
**Ke Lu,** *Inst. of Metal Res., CAS*
**Christian Lüscher,** *U. of Geneva*
**Jean Lynch-Stieglitz,** *Georgia Inst. of Tech.*
**David Lyons,** *U. of Edinburgh*
**Fabienne Mackay,** *QIMR Berghofer*
**Anne Magurran,** *U. of St. Andrews*
**Asifa Majid,** *U. of York*
**Oscar Marín,** *King's Coll. London*
**Charles Marshall,** *UC Berkeley*
**Christopher Marx,** *U. of Idaho*
**David Masopust,** *U. of Minnesota*
**Geraldine Masson,** *CNRS*
**Jason Matheny,** *Georgetown U.*
**Heidi McBride,** *McGill U.*
**C. Robertson McClung,** *Dartmouth*
**Rodrigo Medellín,** *U. Nacional Autónoma de México*
**C. Jessica Metcalf,** *Princeton U.*
**Baoxia Mi,** *UC Berkeley*
**Tom Misteli,** *NCI, NIH*
**Alison Motsinger-Reif,** *NIEHS, NIH* (**S**)
**Suresh Naidu,** *Columbia U.*
**Danielle Navarro,** *U. of New South Wales*
**Daniel Nettle,** *Newcastle U.*
**Daniel Neumark,** *UC Berkeley*
**Beatriz Noheda,** *U. of Groningen*
**Helga Nowotny,** *Vienna Sci. & Tech. Fund*
**Rachel O'Reilly,** *U. of Birmingham*
**Pilar Ossorio,** *U. of Wisconsin*
**Andrew Oswald,** *U. of Warwick*
**Isabella Pagano,** *Istituto Nazionale di Astrofisica*
**Elizabeth Levy Paluck,** *Princeton U.*
**Jane Parker,** *Max Planck Inst. Cologne*
**Giovanni Parmigiani,** *Dana-Farber Cancer Inst.* (**S**)
**Daniel Pauly,** *U. of British Columbia*
**Ana Pêgo,** *U. do Porto*
**Samuel Pfaff,** *Salk Inst.*
**Julie Pfeiffer,** *UT Southwestern Med. Ctr.*
**Philip Phillips,** *UIUC*
**Matthieu Piel,** *Inst. Curie*
**Kathrin Plath,** *UCLA*
**Martin Plenio,** *Ulm U.*
**Katherine Pollard,** *UCSF*
**Elvira Poloczanska,** *Alfred-Wegener-Inst.*
**Julia Pongratz,** *Ludwig Maximilians U.*
**Philippe Poulin,** *CNRS*
**Jonathan Pritchard,** *Stanford U.*
**Lei Stanley Qi,** *Stanford U.*
**Trevor Robbins,** *U. of Cambridge*
**Joeri Rogelj,** *Imperial Coll. London*
**Amy Rosenzweig,** *Northwestern U.*
**Mike Ryan,** *UT Austin*
**Miquel Salmeron,** *Lawrence Berkeley Nat. Lab*
**Nitin Samarth,** *Penn State U.*
**Erica Ollmann Saphire,** *La Jolla Inst.*
**Joachim Saur,** *U. zu Köln*
**Alexander Schier,** *Harvard U.*
**Wolfram Schlenker,** *Columbia U.*
**Susannah Scott,** *UC Santa Barbara*
**Anuj Shah,** *U. of Chicago*
**Vladimir Shalaev,** *Purdue U.*
**Jie Shan,** *Cornell U.*
**Beth Shapiro,** *UC Santa Cruz*
**Jay Shendure,** *U. of Wash.*
**Steve Sherwood,** *U. of New South Wales*
**Brian Shoichet,** *UCSF*
**Robert Siliciano,** *JHU School of Med.*
**Lucia Sivilotti,** *U. Coll. London*
**Alison Smith,** *John Innes Ctr.*
**Richard Smith,** *UNC* (**S**)
**Mark Smyth,** *QIMR Berghofer*
**John Speakman,** *U. of Aberdeen*
**Tara Spires-Jones,** *U. of Edinburgh*
**Allan C. Spradling,** *Carnegie Institution for Sci.*
**V. S. Subrahmanian,** *Northwestern U.*
**Ira Tabas,** *Columbia U.*
**Eriko Takano,** *U. of Manchester*
**Patrick Tan,** *Duke-NUS Med. School*
**Sarah Teichmann,** *Wellcome Sanger Inst.*
**Rocio Titiunik,** *Princeton U.*
**Shubha Tole,** *Tata Inst. of Fundamental Res.*
**Maria-Elena Torres Padilla,** *Helmholtz Zentrum München*
**Kimani Toussaint,** *Brown U.*
**Barbara Treutlein,** *ETH Zürich*
**Jason Tylianakis,** *U. of Canterbury*
**Wim van der Putten,** *Netherlands Inst. of Ecology*
**Ivo Vankelecom,** *KU Leuven*
**Henrique Veiga-Fernandes,** *Champalimaud Fdn.*
**Reinhilde Veugelers,** *KU Leuven*
**Bert Vogelstein,** *Johns Hopkins U.*
**Julia Von Blume,** *Yale School of Med.*
**David Wallach,** *Weizmann Inst.*
**Jane-Ling Wang,** *UC Davis* (**S**)
**Jessica Ware,** *Amer. Mus. of Natural Hist.*
**David Waxman,** *Fudan U.*
**Chris Wikle,** *U. of Missouri* (**S**)
**Terrie Williams,** *UC Santa Cruz*
**Ian A. Wilson,** *Scripps Res.* (**S**)
**Hao Wu,** *Harvard U.*
**Wei Xie,** *Tsinghua U.*
**Yu Xie,** *Princeton U.*
**Jan Zaanen,** *Leiden U.*
**Kenneth Zaret,** *UPenn School of Med.*
**Bing Zhu,** *Inst. of Biophysics, CAS*
**Xiaowei Zhuang,** *Harvard U.*
**Maria Zuber,** *MIT*

EDITORIAL

# COVID-19 vaccines for children

Earlier this month, the US Centers for Disease Control and Prevention (CDC) recommended Pfizer's COVID-19 messenger RNA (mRNA) vaccine for children between 5 and 11 years of age—that's 28 million children. Yet surveys show that 42 to 66% of parents of these children are reluctant or opposed to seeking this protection. Without vaccination, it is likely that almost everyone—including young children—will be infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) at some point in their lives. So, the question for parents and caregivers is: Which is worse, vaccination or natural infection?

Make no mistake—COVID-19 is a childhood illness. When SARS-CoV-2 entered the United States early in 2020, children accounted for fewer than 3% of cases; today, they account for more than 25%. More than 6 million US children have been infected with SARS-CoV-2, including 2 million between the ages of 5 and 11. At the end of October 2021, about 100,000 children per week were infected. Of the tens of thousands of children who have been hospitalized, about one-third had no preexisting medical conditions, and many have required the intensive care unit. Almost 700 children have died from COVID-19, placing SARS-CoV-2 infection among the top 10 causes of death in US children. No children have died from vaccination.

Many parents are concerned that Pfizer's mRNA vaccine was not adequately tested in young children. In a study of approximately 2400 children between 5 and 11 years of age, performed when the Delta variant was the dominant strain, vaccine efficacy was 90.7% against symptomatic disease. However, Pfizer's phase 3 study in adults involved about 40,000 participants. How could the CDC be certain that the vaccine was safe in children given the small size of the study, specifically regarding the problem of myocarditis? In postauthorization studies, myocarditis occurred in approximately 5 per 1,000,000 individuals receiving mRNA COVID-19 vaccines, possibly as high as 1 per 10,000 in young men. But context is important. Vaccine-associated myocarditis has been relatively mild and self-limited—an outcome fundamentally different from the cardiac effects associated with acute COVID-19 or multisystem inflammatory syndrome, which typically involve cardiac dysfunction and require critical care. Moreover, both in Israel and the United States, the incidence of myocarditis in children 12 to 15 years of age receiving mRNA vaccines is less than that in the 16- to 25-year-old age group. And because the dose of Pfizer's mRNA is one-third that given to older adolescents, myocarditis in the younger age group will likely be even rarer.

Myocarditis is only one piece of the risk-benefit analysis. Children need to go to school, play with friends, and participate in extracurricular activities for their social and emotional development. This is their life. Since August 2021, more than 2000 schools in the US have been forced to close because of COVID-19 outbreaks, affecting more than 1 million students. The disruption of school activities has harmed children more than any detectable vaccine side effect, including worsening of mental health, widening education gaps, and decreased physical activity. These harms have disproportionately affected people of color, Indigenous persons, and individuals of lower socioeconomic status, further exacerbating inequities. Avoidance of routine health care and routine vaccination has also emerged, with potentially devastating future consequences. Furthermore, children live closely with and rely on adults to whom they can pass SARS-CoV-2 infection—adults who can be overwhelmed by this infection. And children grow up. Countries are going to need a highly protected population for as long as COVID-19 exists in the world, which will likely be for years if not decades. Vaccinating all children against SARS-CoV-2 could be among the most impactful public health efforts the US has seen in decades.

> "...a choice not to get a vaccine is not a risk-free choice; rather, it's a choice to take a different and more serious risk."

Although it is true that most children experience asymptomatic or mild disease, some will get quite sick, and a small number will die. It's why children are vaccinated against influenza, meningitis, chickenpox, and hepatitis—none of which, even before vaccines were available, killed as many as SARS-CoV-2 per year.

Some parents are understandably hesitant to vaccinate their young children. However, a choice not to get a vaccine is not a risk-free choice; rather, it's a choice to take a different and more serious risk. The biomedical community must strive to make this clear to the public. It could be one of the most important health decisions a parent will make.

**–Jeffrey S. Gerber* and Paul A. Offit**

**Jeffrey S. Gerber** is an associate professor of Pediatrics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA and deputy director of the Center for Pediatric Clinical Effectiveness at Children's Hospital of Philadelphia, Philadelphia, PA, USA. gerberj@chop.edu

**Paul A. Offit** is the director of the Vaccine Education Center in the Division of Infectious Diseases at Children's Hospital of Philadelphia, Philadelphia, PA, USA and a professor in the Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. offit@chop.edu

---

*J.S.G. is site principal investigator for the Moderna–National Institutes of Health KidCOVE trial at Children's Hospital of Philadelphia.

10.1126/science.abn2566

Environmental Science major Giannina Ramirez in William Gilhooly's laboratory at Indiana University-Purdue University Indianapolis (IUPUI).

PHOTO: PROVIDED BY LSAMP

# LSAMP-NICE faculty awards let STEM students benefit from international research

*Science is a collaborative activity, whether it's sharing resources, filling in the expertise gap, benefiting from the comradery of colleagues and getting another—perhaps foreign—perspective on a project, or training the next generation of researchers. Students, particularly those from underrepresented groups, also benefit from these interactions—by gaining new skill sets, by enhancing their resumes, and especially by seeing scientists who look like them and following in their footsteps. Faculty awards given to the Louis Stokes Regional National Science Foundation (NSF) Center of Excellence (LSAMP-NICE) aim to make that happen by bringing the world to both faculty and their underrepresented science, technology, engineering, and mathematics (STEM) students.*

Many picture scientists bound to their bench doing labwork. Yet even STEM scientists are more mobile than is typically thought—often traveling to colleagues' research institutions and even engaging in field work.

Associate professor of earth sciences William Gilhooly has some long-term collaborations, including those with geologists from the University of Campinas and the University of São Paolo (USP) in the State of São Paulo, Brazil. He has hosted a couple of Brazilian researchers—a graduate student and a postdoc—in his Indiana University-Purdue University Indianapolis (IUPUI) lab and was awarded a CAPES-PrInt international (Coordination of Improvement of Higher Education Personnel Program for Institutional Internationalization) exchange fellowship through the Brazilian Ministry of Education to teach a graduate course at USP.

"As far as the student experience, I am a big believer in the opportunities [afforded by] doing international research. One of the things that I really like about my profession, my career, is getting to work with international colleagues from all over the world, especially when you actually get to go and work in other countries," Gilhooly says.

"Whether it's at universities or going to remote field locations, you can really experience life in the culture in ways you wouldn't if you [just] went to all the big cities as a tourist." It's an amazing opportunity for students from different cultures to be able to meet and connect with other young people at a similar stage, age, and academic range as themselves, but with a different life experience growing up in a different country, he adds.

Gilhooly is currently working with an IUPUI environmental science student with Hispanic heritage, Giannina Ramirez, who was looking for more opportunities to work with the lab. He applied for, and was awarded, an LSAMP-NICE faculty award that will

Ngäbe artisan Dilma Jiron shows Orlando Pelcastre, a Heritage University environmental science student visiting Panama, how to break down fibers to make cordage using traditional methods.

allow his student to do some scientific research overseas while enabling him to kickstart a more formal student exchange between universities in the two countries.

## Faculty award is for the students

Gilhooly's award "is a faculty award, but the application and portfolio [that] the faculty submits are based on the work they do with students, as a part of training the students to do scientific research," says Denise Yates, a senior research scientist at the University of Texas at El Paso and a co-principal investigator (co-PI) on the LSAMP-NICE NSF grant. <<*see sidebar on next page*>>

The application materials put it this way: "An awarded faculty mentor would receive $8,000 to support the development and advancement of international research experiences for students."

Important as that is, it's not just about the scientific research. The "international" and "experience" aspects play at least as important a role.

"We want the faculty to take advantage of their (sometimes) existing collaborations with international faculty, and to highlight this for their student mentees," Yates says. At the same time, they want to give underrepresented minorities in STEM the chance "not only to participate in the experience, but to find firsthand why it's important that they be provided an opportunity to expand their horizons and to enrich their portfolios," she adds.

And the benefits go far beyond just those for students who want to travel, especially for those are who are black, indigenous, or people of color (BIPOC). Higher education is still a largely white institution, at least in the STEM fields. There are very few BIPOC in faculty positions to serve as role models, explains Gilhooly. "That makes it really difficult to recruit or retain people, because there are very few people who look like them."

"I think about these faculty awards as representative of the types of activity we are sponsoring through the NSF International Center of Excellence," says Zakiya S. Wilson-Kennedy, assistant dean for diversity and inclusion in the College of Science at Louisiana State University, and an LSAMP-NICE co-PI.

Researchers working internationally often have collaborators nearby their research sites, or in some cases (as will be seen below) functioning as integral partners at those sites. Exposing students to these relationships is fundamental to the LSAMP-NICE mission.

## (Not) All I got were some rocks

Gilhooly's lab looks for signs of chemical signals contained in rock, likely created by biological reactions, and uses these biosignatures to better understand what Earth and life on it was like a long time ago. Analyzing ancient photosynthetic activity and oxygen production may offer clues to the evolution of animal life.

For the LSAMP-NICE

Jessica L. Black, director of the Center for Indigenous Health, Culture and Environment and professor of environmental science at Heritage University.

Costa Rica International Research Experience student Nicole Carley of Lac Courte Oreilles Ojibwe College together with her mentor Jessica Black and Ngäbe indigenous experts Celestino Mariano Gallardo (wildlife) and Armando Sire (medicinal plants) as she investigates dispersal vectors of the invasive species *Zingiber spectabile*, commonly known as beehive ginger.

PHOTOS: TOP, DENISE SILFEE; BOTTOM TWO, PROVIDED BY LSAMP

project, he and his mentee will go down to Brazil and collect 2-billion-year-old rock samples in the field with a colleague from USP and her students. They will do some initial analyses at USP before returning to Indiana, where they will continue analyzing splits of the same subsets of the samples. The researchers will keep each other updated regularly through virtual meetings and "then ultimately move toward getting a publication," he explains.

"One of the big things the faculty award will do is to help further strengthen my professional and personal ties to researchers in Brazil," Gilhooly adds. "And so, one of the big goals of this faculty award is to establish the beginnings of a bidirectional student exchange. There are a lot of analytical resources at both of those universities, and a lot of incredibly talented expertise."

## Representation matters

Jessica L. Black, director of the Center for Indigenous Health, Culture, & the Environment at Heritage University (HU) in Toppenish, Washington, is steeped in introducing a STEM education to underrepresented minorities. HU, where she is a professor of environmental science and studies and chair of the Science Department, is a Native American-Serving Nontribal and Hispanic-Serving Institution located on the traditional lands of the Yakama nation, and a long-standing partner of the All Nations LSAMP (ANLSAMP).

Black was given charge of the ANLSAMP Costa Rica International Research Experience (CRIRE) by Steve Dupuis, a native scholar and PI for LSAMP-NICE who was running CRIRE at the time. As she already had a partnership with the Ngäbe and Buglé people nearby in Panama, she asked her colleague Celestino Mariano Gallardo (now president of their traditional government, the Congreso General Ngäbe Buglé) if he and other indigenous experts would come to Costa Rica to work with the visiting Native American students and their mentors in order to guide them in their research projects, "sharing their traditional ecological knowledge in a global indigenous exchange," she explained.

CRIRE is "very much classic scientific research in tropical ecology– [such as] investigating different species of butterflies, or investigating an invasive flower that's popping up in the forests," she says. "The Native American students from the United States all indicated the most impactful part of the program was learning about the natural world from the local indigenous experts. And the Ngäbe really enjoyed it, because they got to see what our students were learning and saw the potential for their own students. They got to experience what scientific research can be. It became clear to them how much they had to offer."

### Louis Stokes Alliances for Minority Participation (LSAMP)

The Louis Stokes Alliances for Minority Participation (LSAMP), currently represented by 57 member alliances across the United States, was authorized by the U.S. Congress in 1991. Its mission is to help diversify the nation's science, technology, engineering, and mathematics (STEM) workforce (including academia) by funding institutions of higher learning to implement evidence-based strategies for recruiting and retaining students historically underrepresented in STEM disciplines, and imbuing them with academic integration, social integration, and professionalism.

In 2018 the Louis Stokes Regional National Science Foundation (NSF) International Center of Excellence (LSAMP-NICE) received an NSF award–as one of six NSF regional centers to do so–to broaden underrepresented students' participation in international collaborative research through mentorship, research experience, internships, conferences, and other programming opportunities with an aim toward social professionalization. Funded through the NSF Broadening Participation initiative, it is a collaboration of four institutions: Salish Kootenai College, a tribal college in Pablo, Montana (Steve Dupuis, LSAMP-NICE PI); Jackson State University, a historically black university in Jackson, Mississippi (Glake Hill and Martha Mondoa-Tchounwou, co-PIs); Louisiana State University, a research-intensive (R1) institution in Baton Rouge, Louisiana (Zakiya S. Wilson-Kennedy, co-PI); and the University of Texas at El Paso, a Hispanic-serving R1 institution (Denise Yates, co-PI).

## Listening rather than telling

Given the success of CRIRE, Black wanted to assess the potential for a similar—but different—program in Panama: It would be more along the lines of "applied community-based research with an indigenous community partner."

Wary of what she saw as a history of "extractive science," Black wanted to "create and implement programs that are built in partnership and reciprocity with indigenous communities." She also wanted programs that will help engage, retain, and enhance the next generation of scholars underrepresented in the STEM fields.

2019 Costa Rica International Research Experience participants at the Las Cruces Biological Station in San Vito, Costa Rica.

Black had been working with the Ngäbe through their traditional government and several years ago did a series of "listening sessions" to assess what they felt were their needs and concerns—a reversal of a traditional power dynamic of telling community partners what they needed. "One of the strongest themes to emerge was access to water and concerns about changes in the hydrology they were observing."

This seems like "a great training opportunity for our students as well as members of the communities we work with," Black says.

Testing protocols have to be established; sources of contamination have to be investigated if any are found. And locals have to be trained to carry out monthly testing while the international exchange program is not in session.

"We want to build a program that is sustainable. There are many Ngäbe undergraduate students in Panama whom we feel would be very interested, and the Panamanian universities themselves are very interested in this kind of partnership," she notes.

So Black applied for, and was granted, an LSAMP-NICE faculty award, in part to facilitate a formal CRIRE-like exchange—which included establishing or renewing memoranda of understanding with Panamanian universities and community groups, and providing for an HU student mentee to accompany Black to Panama. "And there's funding for me to pay for the water-quality sampling for the partner communities."

PHOTO: PROVIDED BY LSAMP

## Ambassadors for science

"Most of our students are first-generation college students, many are from bilingual households, and very few have had the opportunity to travel much outside of our valley," says Black. "We're taking these students into a whole different world—one where their ability to speak Spanish is highly valued and critical to the success of the program. We want them to understand how strong they are, how valuable their skill sets and their leadership qualities are."

Black says she is "focused on diversifying the professoriate by building programs to transform the system so it's more responsive to the far more diverse set of students we have today, and to enhance the next generation of STEM scholars."

"The faculty award shows how much this work is valued, and that it's worthwhile investing in," she explains. "My hope is that I can use this as a pilot to secure additional funds and keep doing this work."

**Sponsored by**

LSAMP-NICE
NSF International Center of Excellence

# NEWS

## IN BRIEF

*Edited by*
**Lila Guterman**

ENVIRONMENT

## Huge marine reserve to be revoked

The Pacific nation of Kiribati, which in 2008 created the Phoenix Islands Protected Area, then the world's biggest marine protected area, announced this week it will abolish the reserve and allow fishing. Officials said the nation did not receive compensation donors promised it for revenue lost after its 2015 ban on fishing in the reserve, which is the size of Japan. "Strictly managed fishing under conservation measures" will now be allowed in the reefs around the islands, as well as industrial fishing, mostly for skipjack tuna for canning, in the rest of the reserve, the fisheries ministry secretary said. Marine scientists decried the decision, saying it would harm coral reef ecosystems and populations of tuna, which spawn in the reserve. Allowing fishing there is "a terrible mistake," said marine scientist Daniel Pauly of the University of British Columbia, Vancouver.

## Pandemic clampdown in Europe

**COVID-19** | Facing record-breaking rates of coronavirus infection, several European countries have imposed new restrictions on businesses and residents, especially for those who are not vaccinated. This week, Austria imposed a new "lockdown" for anyone over the age of 12 who is not fully vaccinated against SARS-CoV-2; they can only leave their homes for work, school, grocery shopping, exercise, medical appointments, or other "important reasons." Those who break the rules face fines of up to €1450. Roughly 65% of the country's population is fully vaccinated. In neighboring Germany, this week several regions imposed rules that only allow vaccinated people, or those recently recovered from COVID-19 and therefore likely naturally immune, into hair salons, restaurants, bars, theaters, and concerts. On 12 November, the Netherlands imposed a 3-week partial lockdown in which shops and restaurants close at 8 p.m. and spectators are banned from sporting events. After lifting all pandemic-related restrictions in September, Denmark this week reintroduced the "corona pass" that allows only the vaccinated, recovered, and those with a recent negative coronavirus test into restaurants, bars, and concerts.

## Psilocybin depression success

**MENTAL HEALTH** | Psilocybin, the psychedelic substance in magic mushrooms, was reported to be effective in the largest trial of it yet for treatment-resistant depression. In an unpublished randomized trial involving 233 people for whom at least two antidepressant drugs had failed, COMPASS Pathways, a mental health care company based in London, found that almost one-third of patients who took a single 25-milligram dose of psilocybin experienced remission 3 weeks later, compared with 7.6% of those in a control group. Almost one-quarter of patients who received 25 milligrams remained in remission after 12 weeks. Most participants experienced mild or moderate adverse effects such as headache and insomnia, although 12 people had severe ones, including suicidal behavior and self-injury. The company expects to start a large, phase 3 trial next year.

## Chaco Canyon drill ban planned

**ENVIRONMENT** | President Joe Biden this week proposed a 20-year ban on drilling within 16 kilometers of Chaco Culture National Historical Park in New Mexico, a sacred site to Indigenous peoples. Efforts under former President Donald Trump to expand oil and gas extraction from federal lands around Chaco Canyon, where Ancestral Puebloans built a network of great houses, kivas, and roads from about 850 C.E. to 1250 C.E., had drawn fierce opposition from Native American tribes and archaeologists. The planned ban would pause that push. The U.S. Bureau of Land Management will now set aside the land for

PHOTO: RAIMON KATAOTAO/EYEEM/GETTY IMAGES

2 years while it conducts an environmental analysis, consults with tribes, and solicits public comment on the ban, which is sure to draw praise from environmental groups and criticism from energy companies.

## Califf gets FDA nod—again

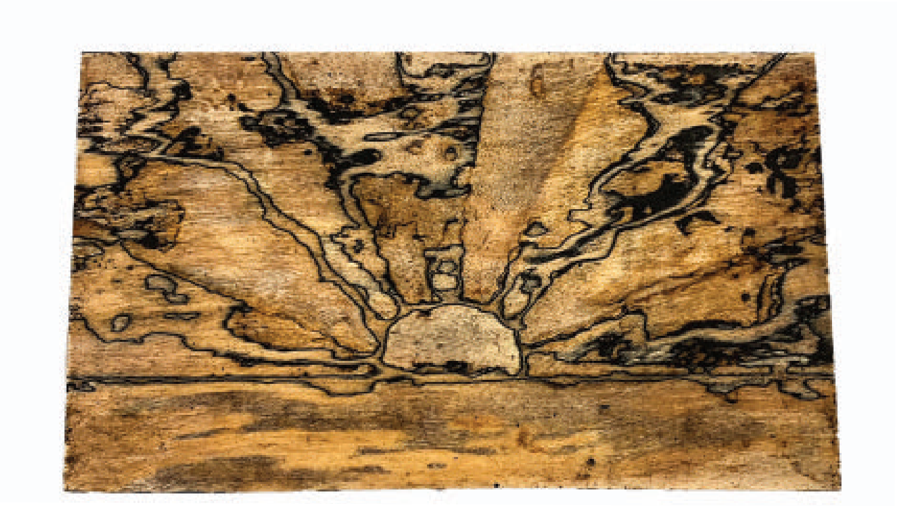
**GOVERNMENT** | Cardiologist Robert Califf, who once led Duke University's clinical research efforts, could get a second stint as head of the U.S. Food and Drug Administration (FDA). President Joe Biden last week nominated Califf to be FDA commissioner, a post he also held for 1 year under former President Barack Obama. Califf will need to be confirmed by the Senate, where several lawmakers have expressed concern about his ties to the pharmaceutical industry.

## Superconductor criticism removed

**PHYSICS** | A critique of one of the highest profile results in physics in recent years has itself come under fire, and the journal that published the critique "temporarily removed" it from its website last week. In September, Jorge Hirsch, a physicist at the University of California, San Diego, published a paper taking aim at a report in *Nature* in October 2020 of the first-ever material to conduct electricity without resistance at room temperature. Hirsch's critique, which appeared in *Physica C: Superconductivity and its Applications*, argued that some of the data in the *Nature* paper looked suspiciously like data in another paper, from 2009. (The authors have rejected those allegations.) In an email to Hirsch, Mary Ma, publisher of *Physica C*, wrote that the editors were concerned that Hirsch published private communication with another scientist and revealed data from other researchers, both without permission. Ma said the paper has been removed until an inquiry is completed.

## Strike creates space debris

**SPACE** | A Russian missile blew up a defunct Soviet radio satellite this week, creating more than 1500 pieces of large debris in low-Earth orbit that threaten other satellites. The high-speed debris cloud, the second largest ever created by military means, forced astronauts and Russian cosmonauts on the International Space Station to temporarily take shelter. Debris from the antisatellite test, the first conducted by Russia, will linger for years before it falls out of orbit, U.S. Space Command said. The United States and India have also shot satellites out of low orbit, creating smaller clouds of space junk that lasted several years. Debris from a 2007 Chinese strike in higher orbit remains a hazard to spacecraft.



MYCOLOGY

## Fungi coaxed to draw

For centuries, woodworkers have prized "spalted" wood, in which fungi create decorative patterns of lines and color. Now, scientists in Switzerland have sped up and controlled the process, creating images such as a sunset (above). By carefully applying strains of fungi to specific kinds of wood and adjusting temperature, acidity, and moisture levels, the researchers can prompt the microbes to draw circles, rectangles, and lines that are curved, straight, or zigzagged. The technique could make otherwise inexpensive and fast-growing types of wood, such as ash and beech, more attractive for furniture and other applications.

PHOTO: EMPA

THREE QS

## COVID-19 vaccine developer looks ahead

Uğur Şahin, the oncologist and immunologist who is CEO of BioNTech, has vaulted to prominence in the past 20 months as his company paired up with Pfizer to produce a powerful messenger RNA (mRNA) vaccine against COVID-19. During his trip last week to a conference in Washington, D.C., *Science* spoke with him about SARS-CoV-2 and other pathogens. His responses have been edited for clarity and brevity.

**Q: Do you think healthy 18- to 64-year-olds need coronavirus vaccine booster shots?**

**A:** The message is very clear: We have absolutely no chance to control the pandemic without providing boosters to everyone. The protection against infection drops to below 50% after 8, 9 months. We now have complete evidence from real-world data that with a booster strategy, Israel managed to get control of the situation. This was not only the elderly. This was the 16-plus population.

**Q: Do you agree with the notion that the virus has likely run through its playbook of significant mutations—and that we're not going to see another huge leap like with the Delta variant? Is that naïve?**

**A:** I would not say naïve. But I would not count on that. You can never deal with the unknown unknown. We need to be careful with gain-of-function mutations. We need to monitor whether the virus will mutate to use different host cell receptors.

**Q: Are you watching any mRNA vaccine candidates for other infectious diseases?**

**A:** There aren't so many mRNA companies at the moment. Moderna has launched several clinical trials based on mRNA candidates. We have encouraging preclinical data for a flu vaccine we are developing with Pfizer. That trial started recently. There are diseases where we believe this should not be a big challenge to get an approved vaccine in a reasonable time because the targets are well known, the mechanism is well known. So you have to just recapitulate previous science with mRNA. There are other infectious diseases, like malaria and TB [tuberculosis], that are not low-hanging fruits.

SCIENCE.ORG/NEWS
Read more news from *Science* online.

GLASGOW LEADERS DECLARATION ON FORESTS & LAND USE
US-CHINA JOINT STATEMENT
PARIS RULEBOOK
MECHANISM
COMMON TIMEFRAMES
ARTICLE 6
KEEP 1.5 ALIVE
HUMAN RIGHTS
WITH NO OVERSHOOT
REAL REDUCTIONS
GLOBAL STOCKTAKE
ADAPTATION FUND
LOSS & DAMAGE NETWORK
GEF
THE END OF THE FOSSIL FUEL ERA
GET THE MONEY TO THE COM
NATURE-BASED SOLUTIONS FOR ADAPTATION & MITIGATION
BIODIVERSITY ECOSYSTEMS
LOCALLY LED ACTION
EQUAL REPRESENTATION
JUST TRANSITION
ACCOUNTABILITY TO VULNERABLE COMMUNITIES
2°C WOULD MEAN THE END 1.5°C WOULD MEAN A FIGHTING CHANCE
HONESTY
OCEAN & CLIMATE DIALOGUE
DON'T BE A FOSSIL FOOL
GLOBAL YOUTH NDS STATEMENT
COY YOUTH STATEMENT
KEEP CARBON UNDERGROUND

A mural at the recent climate summit in Glasgow, U.K., exhorts nations to take steps to limit global warming to 1.5°C.

PHOTO: CHRISTOPH SOEDER/PICTURE ALLIANCE/GETTY IMAGES

CLIMATE CHANGE

# Glasgow pact leaves 1.5°C goal on life support

## Goal of preventing dangerous warming slipping away, but nations boost emissions cuts

*By* **Cathleen O'Grady**

When nations gathered in Glasgow, U.K., earlier this month for the major U.N. climate summit, "Keep 1.5 alive" was a ubiquitous mantra, repeated on protest banners and in speeches by politicians. By the end of the meeting last week, however, it was clear that the international effort to limit global warming to 1.5°C above preindustrial levels—the stretch goal set by the 2015 Paris agreement—is on life support. The Glasgow deal reached on 13 November takes important steps toward limiting warming and "still leaves the door open" to achieving the 1.5°C goal, says Sonia Seneviratne, a climate scientist at ETH Zürich. "But only slightly."

Current commitments to curb greenhouse gas emissions—including new national pledges finalized at the summit—are more ambitious than ever, but still not enough to keep warming below 1.5°C, researchers say. Many are vague and promise results only decades from now, and the pact includes no enforcement mechanisms. Nor does it guarantee that rich nations will provide the financial help that poorer countries will need to control their emissions and cope with climate impacts.

The outcomes were disappointing, although not unexpected, for many developing and island nations that fear warming greater than 1.5°C will bring devastating harm. Still, many attendees were encouraged by signs of progress. More than 100 nations made new pledges to curb emissions that, if realized, could for the first time keep warming below 2°C. And 197 countries agreed to a host of new measures—including standards for reporting national emissions and ground rules for trading credits for carbon removal—that could promote more ambitious pledges and reveal whether countries are actually following through. Nations also agreed to return to next year's summit, to be held in Egypt, with even stronger commitments, aimed at meeting the 1.5°C goal.

The agreement signals a "major shift" in global climate talks, from designing policy frameworks to implementing them, says Rueanna Haynes, an international climate law specialist at the Berlin-based nonprofit Climate Analytics. "It took forever, but we're done now. … The focus is going to shift squarely into implementation," says Haynes, who negotiated at the summit on behalf of the Alliance of Small Island States.

The talks also reflected a new level of realism about what it will take to rein in warming, says Maisa Rojas, a climate scientist at the University of Chile. This time, the agreement embraced a 2018 U.N. scientific report that concluded carbon dioxide emissions will need to drop by 45% by 2030 for the world to stay below 1.5°C of warming. "It's great to see the science so clearly stated," Rojas says.

The bad news, however, is that the world is not now on track to achieve such reductions. Only by halving emissions over the next 10 years and then rapidly dropping them to net zero by 2050 can the world gain a 50% chance of staying below 1.5°C, says Joeri Rogelj, a climate scientist at Imperial College London.

Such projections underscore the importance of having nations revisit and ramp up their pledges next year in Egypt, rather than in the 5 years specified in the Paris pact, commentators say. If countries waited that long to increase their ambitions, reaching the 1.5°C goal would become "basically impossible," Seneviratne says.

Despite the difficulty of achieving the 1.5°C target, many experts were pleased the Glasgow pact retained language reinforcing its importance. Island states threatened by sea level rise—which led the charge in securing the 1.5°C target in the Paris pact—argued the goal is, for them, existential.

Still, Rogelj worries about what will happen if the 1.5°C target is declared dead. "That implies that when we miss 1.5°C, the next target is 2°C," he says, rather than some number in between. But, "Every fraction of a degree matters," says Nerilie Abram, a climate scientist at Australian National University. She notes that even small increases in temperature boost the risks posed by sea level rise and climate shifts. The lower goal, she and others say, should help maintain political pressure for stronger action, even if—or when—it is breached.

Developing nations did not get one big thing they wanted in Glasgow: a new "loss and damage" fund. Fund advocates argued that developed nations, having produced the vast majority of historic emissions, should help developing countries cope with the costs of climate-related extreme events, such as droughts. In the end, the pact promised only a "dialogue" on loss and damage. The issue is expected to be high on the agenda of next year's summit, but, for the moment, "Developing countries are very angry," says economist Sarobidy Rakotonarivo of the University of Antananarivo.

Richer nations did make some, often vague, pledges to help developing countries reduce their emissions and build the infrastructure needed to adapt to climate impacts such rising seas. But researchers say targeting that money effectively will also require funding more science, including higher resolution modeling of climate impacts and studies of the economic impacts of climate-driven extreme events.

Despite Glasgow's mixed results, momentum for meaningful climate action is growing, believes Rebecca Willis, an environmental social scientist at Lancaster University. There is a limit to what summits can achieve, she says, because nearly 200 countries must agree to the outcomes. But she notes other players, including the energy and finance industries, keep a careful watch, and this year's summit left them with plenty to think about. The idea of keeping fossil fuels in the ground, for example, "has gone from being a campaign slogan to being on the desk of every investment banker," Willis says. And although the progress made in Glasgow "doesn't go far enough, it's not fast enough … it's unrecognizable compared to where we were 10 years ago." ■

---

ATMOSPHERIC CHEMISTRY

# High-flying wildfire smoke poses potential threat to ozone layer

## Record Arctic ozone loss linked to Siberian wildfires

By **Paul Voosen**

Two years ago, the crew of the *Polarstern*, a German icebreaker frozen into Arctic sea ice, shot a green laser up into the night. The beam's reflected light was meant to help researchers study icy winter clouds. Instead, the beam encountered something unexpected: a kilometers-thick layer of particles in the stratosphere, more than 7 kilometers up. The haze, the researchers later concluded, was smoke from enormous wildfires that had ripped through Siberia that summer.

The smoke was more than a curiosity. By March 2020, as the Siberian smoke lingered, satellite measurements of ozone levels in the Arctic hit a record low—not quite a "hole," by Antarctic standards, but worryingly low. Although the case is far from closed, it seems likely the smoke helped deplete the ozone, says Kevin Ohneiser, a graduate student at the Leibniz Institute for Tropospheric Research (TROPOS). Similar dips have occurred the past 2 years in Antarctica following Australia's record-breaking "Black Summer" fires, which injected more than 1 million tons of smoke into the stratosphere. "We cannot prove this," he says. "But these [results] seem to be a hint."

The findings, which Ohneiser and his colleagues published last month in *Atmospheric Chemistry and Physics*, suggest climate change may have an unexpected impact on atmospheric chemistry, as smoke from increasingly severe wildfires invades the stratosphere and potentially erodes the ozone layer that screens out damaging ultraviolet (UV) radiation. "Until recently, smoke was really discounted in terms of a global impact," says Catherine Wilka, a stratospheric chemist at Stanford University. Now, she adds, it's "shaping up to be one of the new frontiers."

"This is really new," adds Omar Torres, a remote sensing scientist at NASA's Goddard Space Flight Center. Since the late 1970s, satellites have been capable of tracking smoke particles, easily visible from space because they are strong absorbers of UV light. Until 2017, however, the satellites saw no sign of smoke penetrating the stratosphere in any appreciable amount, Torres says.

The Arctic smoke event is particularly worrisome because it had no business being there. "Everyone thought the Arctic would be really clean," Ohneiser says, because it lacks the thunderstorms that can propel pollutants into the stratosphere, a calm, isolated layer above the troposphere. Today's fiercest wildfires, such as those in Australia, can generate their own towering storm systems, capable of pumping material into the stratosphere like volcanoes. But while Siberia burned, it was trapped in a heat wave and a high-pressure system that smothered

A 2019 heat wave sparked widespread fires in Siberia. The smoke may have wound up in the Arctic stratosphere.

PHOTO: DONAT SOROKIN/TASS/GETTY IMAGES

the convective updrafts that form large storms. The smoke must have had another route to the stratosphere.

In a model not yet published, the TROPOS group attempts to explain how the region could feed smoke so high, invoking a decade-old theory called "self-lifting." Their model suggests the dark smoke particles absorbed sunlight so effectively that they rapidly heated the air around them, causing the smoke to rise. After only a few days, the process could have lofted smoke 10 kilometers above the ground, where winds could then usher it into the low Arctic stratosphere. And indeed, on passes over the Siberian fires, NASA's Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) laser satellite captured plumes of what seemed to be smoke rising from 4 to 10 kilometers, Ohneiser says.

The self-lifting idea, never documented in the troposphere, is controversial. In the small world of fire storm, or pyrocumulonimbus (pyroCB), research, "Somehow the idea has been advanced that the only way smoke aerosol can get to the stratosphere is due to direct injection," says Torres, who identified self-lifting as part of the way that smoke from 2017 fires in British Columbia reached the stratosphere. "But the observations are showing it is still happening when we have no pyroCBs."

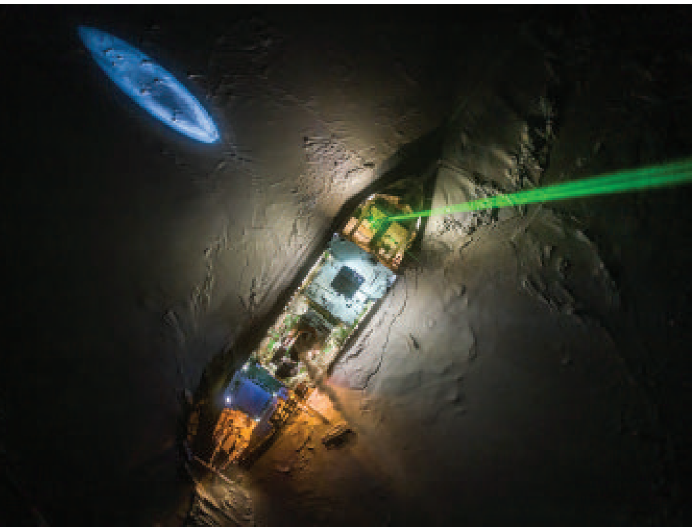
Others are not convinced. Michael Fromm, a pyroCB researcher at the U.S. Naval Research Laboratory, calls it an "extraordinary claim," requiring more robust evidence. He thinks that without the extra boost from a firestorm, smoke is unlikely to penetrate the tropopause, a boundary that helps isolate the stratosphere. Instead of smoke, Fromm believes most of the Arctic particles are lingering sulfate aerosols from Raikoke, a volcano southwest of Russia's Kamchatka Peninsula that in 2019 heaved gas and ash into the stratosphere. He points out that CALIPSO can't distinguish between smoke and sulfates.

But Ohneiser and his colleagues are standing firm. Their advanced lidar measures light absorption and reflection at two different wavelengths, and observations of the Australian fires using the same instrument showed smoke particles have a distinctive signature. These are "unambiguous optical fingerprints of wildfire smoke," Ohneiser says. "There is no room for other interpretations." In the paper, the TROPOS team does see sulfate particles from Raikoke, but they form a thin layer even higher up in the stratosphere.

Once smoke is in the stratosphere, "the potential is certainly there" for it to deplete ozone, says Jessica Smith, an atmospheric chemist at Harvard University. Polar ozone loss depends on chlorine, still lingering in the stratosphere from chlorofluorocarbons and other pollutants even though they were banned decades ago. The chlorine attacks in winter, when thin iridescent clouds form in the stratosphere. Their droplets provide a surface for chemical reactions that result in free radicals of chlorine, which chew through ozone. Smith says smoke particles might boost ozone loss by seeding the formation of these clouds and endowing them with smaller, more abundant droplets.

Smoke particles might also be coated in chemicals such as sulfates that could reduce ozone by directly reacting with chlorine. Or the smoke could somehow strengthen a collar of stratospheric winds called the polar vortex, further chilling the poles and boosting depletion. The loss mechanisms are speculative, Smith says, but they "could take a strong year and tip into an extreme year."

The influence of stratospheric smoke isn't necessarily limited to the poles. At midlatitudes, the stratosphere is much higher, and in theory more insulated from pollution. But as wildfires worsen, Wilka says, smoke might even have a shot at reducing ozone above the midlatitudes, home to most of the world's population, much as the 1991 volcanic eruption from Mount Pinatubo did. Throw enough smoke and other particles up there, she says, and "you can absolutely start driving this chemistry." ■



In 2019, the *Polarstern*, a German ship locked in Arctic ice, shot a laser into the stratosphere. It revealed a thick haze thought to be smoke.

PHOTO: ESTHER HORVATH AND JAKOB STARK/CC-BY

---

**MARINE CONSERVATION**

# Mediterranean mussel faces 'imminent extinction'

## After devastating epidemic, Europe launches new efforts to save one of the world's largest bivalves

*By* **Erik Stokstad**

In 2017, Maite Vázquez-Luis and her colleagues searched the seagrass meadows in Cabrera Archipelago Maritime-Terrestrial National Park, near the Spanish island of Mallorca, for one of the reserve's best known inhabitants: a mussel the size of a skateboard. The fan mussel, an iconic Mediterranean Sea species, can grow 1.2 meters tall and live some 50 years. But the survey found not a single living specimen. "We were shocked," recalls Vázquez-Luis, a marine ecologist at the Spanish Institute of Oceanography. "We couldn't understand what was going on."

Studies soon revealed the animals had succumbed to a mysterious new parasite that may have arrived via ships. Since then, the parasite, perhaps abetted by other microbes, has erased fan mussel populations throughout the Mediterranean, raising fears that the storied mussel could be facing extinction.

Now, two international consortia, to be launched next week, aim to save it. With €5.2 million over 3 years from the European Union and partners, the teams plan to search for scattered survivors that might be resistant to the parasite and figure out how to breed the mussels in captivity. "The situation is really critical," says Jose Rafael Garcia-March, a marine biologist at the Catholic University of Valencia who is coordinating one of the groups. (The other is led by the Ligurian Regional Environmental Protection Agency in Italy.) "We are running against the clock."

The fan mussel (*Pinna nobilis*), also called the noble pen shell, has been highly valued since antiquity. Artisans carved buttons from its pearly shell and the Phoenicians wove luxury garments from the soft,

Before a mysterious disease struck, fan mussels were an attraction for scuba divers, such as these off the coast of Kaş, Turkey, in 2018.

iridescent threads the mussel uses to anchor itself to rocks. In the eastern Mediterranean its flesh is still served, illegally, as a delicacy. Ecologists prize it, too. Individuals filter thousands of liters of seawater every day, helping improve water quality, which is crucial for seagrass meadows. Hundreds of other species live on or inside the shells.

In the 1990s, after overharvesting, pollution, and other threats caused populations to decline, the species won legal protections and began to recover in many places. But the new disease has darkened the outlook. Since it was first spotted in September 2016, it has eliminated every monitored population off Spain, France, Italy, and Greece, more than 200 in all. In 2019, the International Union for Conservation of Nature declared the species critically endangered. “This mass mortality event is one of the biggest things I’ve ever dealt with,” says Francesca Carella, a veterinary pathologist at the University of Naples Federico II.

Researchers have found a new protozoan, *Haplosporidium pinnae*, in dead and dying mussels. The parasite infects connective tissue and produces spores inside digestive glands, eventually killing the bivalves. Sick mussels, too weak to close their shells, become easy picking for predators. They may also become vulnerable to infection by other pathogens such as *Vibrio* and *Mycobacteria*. Some researchers suspect ships brought the parasite to the Mediterranean in their ballast water and that climate change is aiding its spread.

As the outbreak widened, researchers scrambled to protect the mussels. Garcia-March and colleagues moved 215 of them to laboratory tanks and found that cold temperatures kept the parasite at bay. But the chilly mussels wouldn’t reproduce, and when they were warmed, the parasite promptly killed them all. The same pattern played out at an aquarium in Croatia.

Since then, researchers have learned how to induce the mussels to sexually mature, and the new projects aim to figure out how to keep larvae alive to maturity. Researchers at the Paul Ricard Oceanographic Institute and the University of Sassari will also map the genetic diversity of populations to guide restoration, with some trying to identify mussels that might spawn resistant offspring. Patricia Prado, a marine ecologist at the Institute of Agrifood Research and Technology, hopes other genetic research will offer insight into why some populations are reproducing infrequently.

Already researchers and volunteer scuba divers have found some apparently hardy mussels. The discoveries are rare enough that those survivors can make local headlines, and human allies build cages to protect them from predators or damage. The new consortia hope there are more, perhaps holding out in little-surveyed areas—such as the Mediterranean’s southern coast—and in deeper, colder waters. Monitoring suggests at least some mussels are still reproducing off the French coast and near Algeria.

“There is an urgent need to know where the resistant survivors are,” says Saul Ciriaco, an environmental researcher at World Wide Fund for Nature Italy’s Miramare Marine Protected Area.

These survivors could be moved to tanks for breeding or to waters that appear to deter the parasite, perhaps because of temperature or salinity differences. Transplantation is safe, experts say. In research conducted before the epidemic, for example, marine ecologist Stelios Katsanevakis of the University of the Aegean and colleagues showed high survival rates for transplanted animals (although all 20 in a first trial were illegally harvested within 3 days).

In the meantime, researchers are thinking about how to protect bays and lagoons that still have mussels. Some of these waters once hosted huge beds; an estimated 1.8 million fan mussels, for example, carpeted Spain’s Mar Menor lagoon. Now, just a few thousand animals remain as a result of harvesting, accidents, and pollution. Runoff from farm fields is fueling harmful algal blooms that pose an especially serious threat, researchers say. Earlier this month, Spanish officials announced a €382 million, 5-year initiative to curb the pollution; it includes funds to restore fan mussels.

“It’s very good news,” Vázquez-Luis says. And Garcia-March hopes such efforts and the EU-funded projects will mark a turning point in efforts to save the fan mussel. The grim alternative, Prado says, is “witnessing … an imminent extinction of the species.” ■

PHOTO: HELMUT CORNELI/ALAMY STOCK PHOTO

A tourist from Shanghai exercises in a quarantine hotel in Zhangjiajie, in China's Hunan province.

COVID-19

# 'Zero COVID' is getting harder—but China is sticking with it

## Government aims to increase immunity further before relaxing controls and trying to live with the virus

*By* **Dennis Normile**

As China battles its third major COVID-19 outbreak this year, the country is doubling down on the "zero COVID" strategy of mass testing, contact tracing, and containment that many of its Asian-Pacific neighbors have abandoned.

The flare-up of the highly infectious Delta variant, with more than 1000 cases reported since mid-October in 21 of the country's 34 provinces and regions, has led China to take some extraordinary measures. Beijing residents who had the bad luck of traveling to what became COVID-19 hot spots were simply barred from returning home. On 31 October, after a single confirmed case was linked to Shanghai Disneyland, authorities tested 34,000 visitors to the park before allowing them to leave. Testing took until near midnight and turned up only negatives.

Such logistical challenges, increasing lockdown fatigue, and growing public opposition have led New Zealand, Singapore, and parts of Australia to give up on zero COVID, opening their economies and accepting a surge in cases that is small by international standards yet much bigger than anything they had seen before. But China intends to stay the course until vaccination rates—both at home and abroad—are so high that SARS-CoV-2 can do little damage.

"The current cost of adopting a 'zero-transmission' policy is indeed high, but it would cost even more if we live with [the virus] and open up," Zhong Nanshan, a respiratory specialist at Guangzhou Medical University and China's leading COVID-19 adviser, said in an early November interview with CGTN, a state-owned broadcaster.

China's containment strategy "has been a great success in terms of minimizing the number of COVID-19 infections and deaths," says Benjamin Cowling, an epidemiologist at the University of Hong Kong. China has only reported 127,018 cases of COVID-19 and 5697 deaths, according to the World Health Organization. The United States, with roughly one-quarter of China's population, has seen 46.5 million cases and more than 750,000 deaths. But the success "comes at a great price in other areas," says Dale Fisher, an infectious disease specialist at the National University of Singapore, including the economic and social toll. Those include severe restrictions on international travel: Incoming tourism is barred, entry visas for other purposes are difficult to get, and travel abroad for Chinese is discouraged.

Experts outside the country are divided on whether it makes sense to continue the policy. Evolutionary virologist Edward Holmes of the University of Sydney says China is only putting off the inevitable. "I think China's zero policy is completely unsustainable and ill-conceived for a virus that is obviously going to be endemic," Holmes says. But Nick Wilson, a public health scientist at the University of Otago, Wellington, says sticking to zero COVID "might possibly still be the best strategy even at this stage—if it gives China time to increase vaccination coverage, including to children, deliver more booster doses, and allow time for improved vaccines to be developed."

The increasing supply of vaccines, a desire to open borders, and the difficulty of stamping out Delta outbreaks led New Zealand to move from an elimination strategy to "suppression," says Michael Baker, also a public health specialist at Otago. But the shift "involves complex trade-offs and considerable uncertainty," he says.

Currently, more than 75% of Chinese adults are fully vaccinated, and even children as young as age 3 are now getting the shots, according to Caixin Global, a China-based news website. That is not nearly enough to keep the Delta variant at bay; even countries with vaccination rates higher than 85%, such as Singapore, are reporting substantial numbers of new infections.

George Gao, director of the Chinese Center for Disease Control and Prevention, predicts China will achieve 85% coverage early next year, but says "we are working very hard" to push the rate even higher. Booster shots will be rolled out in the first half of next year, Zhong told CGTN, which he said could help China reach herd immunity, a milestone that has proved elusive elsewhere. After that, China could relax its travel restrictions for countries with high vaccination rates, Zhong said—possibly by the end of 2022.

But the quarantine measures for visitors will likely stay in place, Cowling says. They are getting even stricter. Guangzhou, a booming metropolis near Hong Kong, just completed a 5000-room facility where incoming travelers—even the vaccinated—will have to spend 14 days of quarantine; previously they stayed at hotels scattered through the region. Other cities are building or planning similar facilities. Overseas travelers arriving in the northern city of Shenyang now face 1 month of hotel quarantine followed by another month of home confinement. In some cities, even domestic travelers from high-risk areas face quarantines.

In the end, vaccination may drive COVID-19 mortality down to match that of seasonal flu, which Zhong pegs at 0.1% to 0.5% in China. At that level, "we may be able to live with it," he said. But not now, he cautioned: "China will still stick to the 'zero-infection' policy for a rather long time." ■

With reporting by Bian Huihui.

PHOTO: CHEN SIHAN/XINHUA/GETTY IMAGES

INFECTIOUS DISEASE

# To fight Lyme disease, vaccine targets ticks

## Ticks fed poorly, transmitted pathogen less often to guinea pigs given mRNA vaccine

*By* **Meredith Wadman**

Erol Fikrig had spent 10 years pursuing a vaccine that would take a new approach to protecting people from Lyme disease, a growing bane in the United States: He wanted to target not the pathogen, but the tick that transmits it. Then, at a June 2019 meeting in Killarney, Ireland, he heard immunologist Drew Weissman of the University of Pennsylvania describe what was then a little-known technology: messenger RNA (mRNA) vaccines. In a flash, Fikrig saw a way forward. The Yale School of Medicine infectious disease physician collared Weissman and asked whether the technology might work against the deer tick that transmits Lyme disease in the United States. "I'd like to pursue that," Fikrig recalls Weissman saying.

MRNA technology is now famous for delivering vaccines against COVID-19, and this week it achieved another distinction with an experimental Lyme preventive announced by the collaboration launched in Ireland. "It's the first vaccine [intended for humans] against an infectious disease that does not target the pathogen," Fikrig says. The mRNA vaccine, administered to guinea pigs, turned tick bites red and inflamed. The ticks fed poorly, fell off early, and often failed to transmit the Lyme-causing bacterium. Researchers hope the vaccine will one day work the same way in humans.

Each year the deer tick (*Ixodes scapularis*) transmits Lyme disease to as many as half a million people in the United States.

It's "a beautiful study," says Ruth Montgomery, a cellular immunologist at Yale who was not involved with the work. "Potentially a mechanism like this could be very important in a number of tick-borne diseases."

Others are impressed by the team's technological feat. The researchers packed 19 distinct mRNA snippets, each encoding a protein, or antigen, from deer tick saliva, into a single vaccine; COVID-19 mRNA vaccines deliver just one. "The mRNA vaccine saved us from COVID for sure," says microbiologist Jorge Benach of Stony Brook University, who co-discovered *Borrelia burgdorferi*, the tick-borne spirochete that causes Lyme disease. "Now [Fikrig] is using stunning technology … with more than one antigen simultaneously. … I think it will be very very useful for future vaccines."

In the United States, Lyme disease is the most common infection transmitted from animals to people, with up to 476,000 cases each year. *B. burgdorferi* triggers a flulike illness and trademark skin rash, and can also invade the brain, nerves, heart, and joints, sometimes leading to permanent nerve damage and arthritis. "For some people Lyme disease can cause major problems," says Adriana Marques, who runs Lyme disease trials at the National Institute of Allergy and Infectious Diseases. Antibiotics can treat Lyme disease in its early stages, but a growing number of people—at least 1.6 million in 2020, according to one estimate—now suffer chronic consequences. No human vaccine is currently available, although one is in human trials.

That vaccine targets *B. burgdorferi* itself, but Fikrig thought a vaccine aimed at targeting the deer tick (*Ixodes scapularis*) might stymie the bacteria. Tick saliva secretes agents that help transmit the pathogen, but those proteins are "difficult to make in the lab," Fikrig says. "The beauty of the mRNA vaccine is … you don't need to make the protein—the body does that for you."

In many people, tick bites go unnoticed, allowing ticks to feed uninterrupted. The new vaccine, with its multiple mRNA snippets that order host cells to make important tick salivary proteins, primed the guinea pigs' immune systems to react to a tick bite. Within 18 hours after the ticks attached, most bites were transformed into red, inflamed, and likely itchy wounds, according to work published this week in *Science Translational Medicine*.

That's important because *B. burgdorferi* is rarely transmitted from tick to host before 36 hours into a tick meal that often lasts 4 days or more. And when scientists pulled ticks off soon after the bite site became inflamed—as a human might do—the transmission of *B. burgdorferi* was blocked.

"Everybody should look at these as very, very promising results," says epidemiologist Sam Telford of Tufts University.

Still, much of the protection will probably hinge on whether people notice an itchy, red, tick bite and manage to pull the tick off early. When three infected ticks were attached and allowed to feed on guinea pigs until they were sated, 60% of vaccinated animals became infected, almost as many as control animals. And whether vaccinated people will react to the ticks as the guinea pigs do remains an open question. "Responding to the bite is really, really a cool thing," says Uğur Şahin, CEO of BioNTech, which with Pfizer developed an mRNA vaccine against COVID-19. "We have to see if it holds true when translated into the human situation."

Scientists note that a successful human vaccine, which would need a pharmaceutical company sponsor, could also protect people from other, rarer pathogens also transmitted by the deer tick, such as babesiosis. A vaccine might also induce a reaction to other tick species, such as those that transmit Lyme disease in Europe.

Benach thinks the new vaccine might also have to be loaded with mRNA that targets the pathogen. "I would like to see it supplemented with pathogen [mRNA] because it's the pathogen that … makes you ill," he says.

Telford, who has spent years working in communities with high Lyme disease burdens, hopes the vaccine will offer one solution to a growing problem: "People are fed up with tick bites. People are fed up with Lyme disease. They want something done." ■

PHOTO: LADISLAV KUBEŠ/ISTOCK.COM

FEATURES

# GETTING THE BIG PICTURE OF BIODIVERSITY

## Satellites and other remote sensing tools offer new ways to study ecosystems—and maybe even save them

*By* **Elizabeth Pennisi**

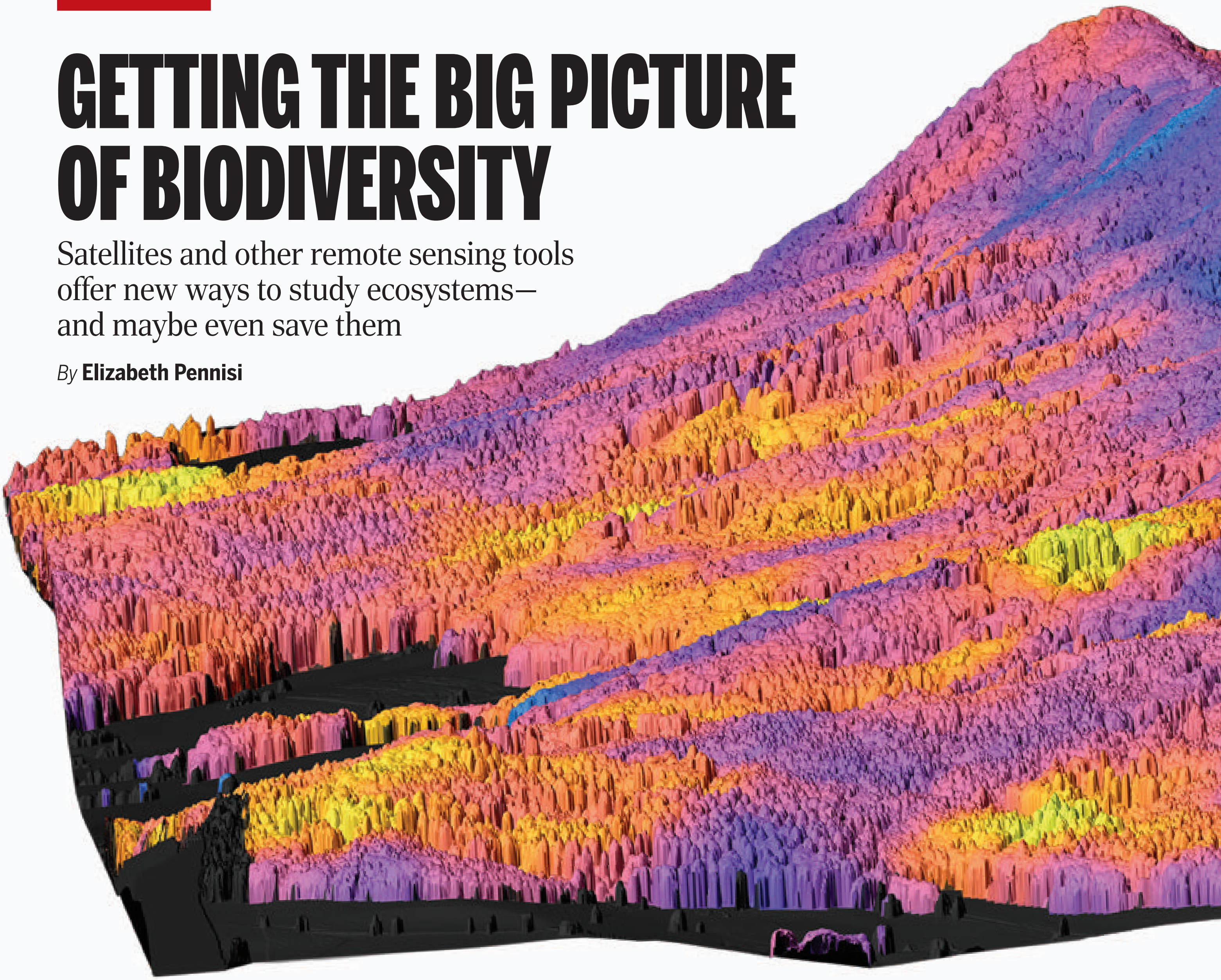


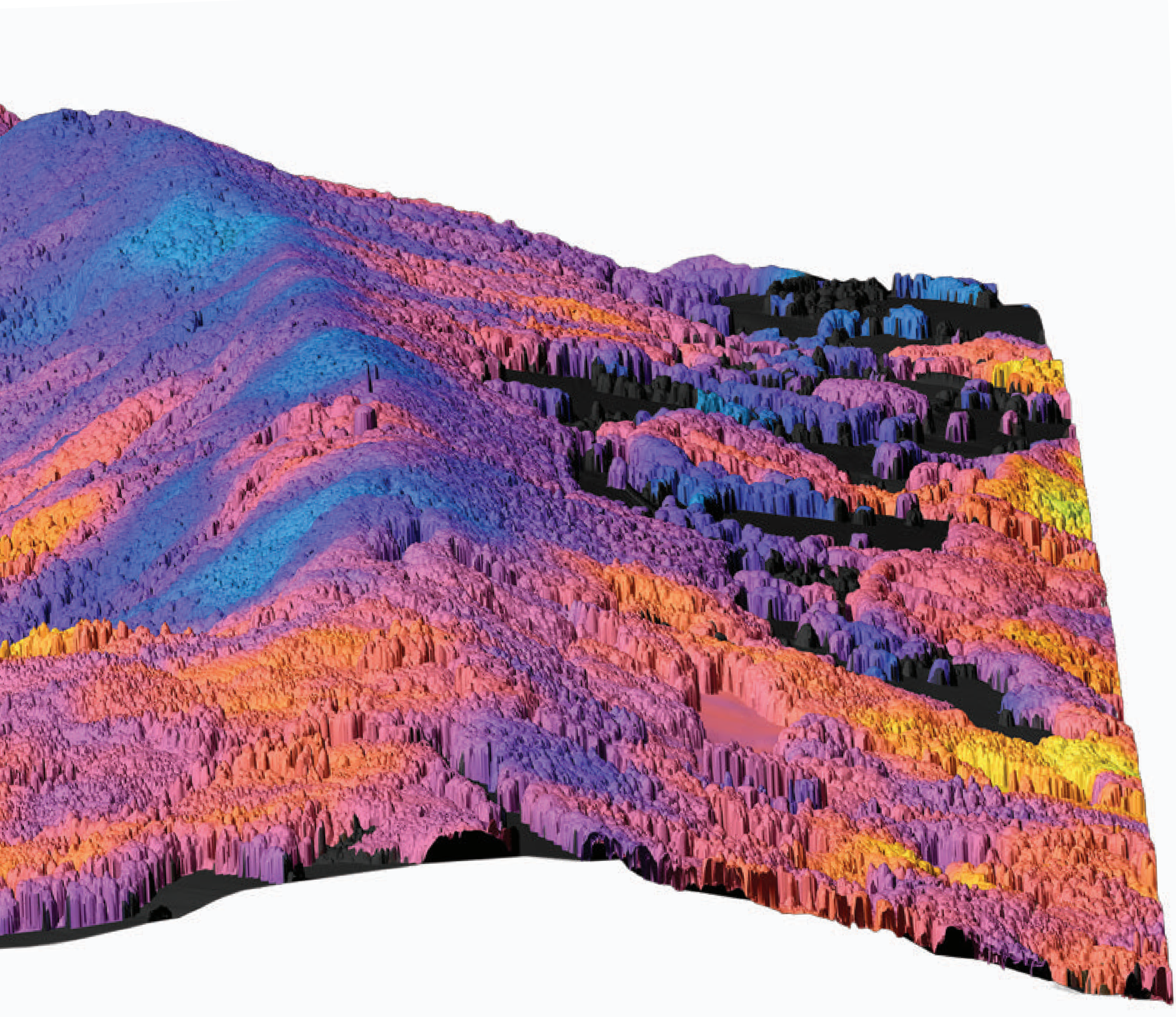
IMAGE: F. MORSDORF, *REMOTE SENSING OF PLANT BIODIVERSITY*. DOI.ORG/10.1007/978-3-030-33157-3_4 (2020)

Jeannine Cavender-Bares began her scientific career with her hands in the dirt. Now, she has her eyes in the sky. As a teenager she helped her biologist dad dig deep into the leaf litter of Florida oak forests to catalog the diversity of slime molds, protists best known for uniting into unsightly blobs that creep across the landscape. But it was the oaks overhead that really fascinated her and became the focus of high school science fair entries and her graduate school research.

In forest plots across the southeastern United States, she planted acorns to study how local conditions affect photosynthesis and growth of different oak species. She froze the seedlings' stems to study how they transport water to leaves and climbed into the forest canopy to measure gases emitted by mature trees.

But such studies could only provide a snapshot of one forest at a time. To get the big picture of forests around the world, Cavender-Bares has sought a higher vantage. Now a plant ecologist at the University of Minnesota, Twin Cities, Cavender-Bares has devised ways to translate light measured by spectrometers flown over forests into insights about their health and resilience. She and others have found this light, captured from an airplane or satellite, holds clues to intimate details such as photosynthesis levels, the genetic diversity of the trees, and even the microbial inhabitants of the soil they grow in.

Such remote sensing methods are not only revolutionizing how scientists such as Cavender-Bares study ecosystems, they're also poised to become powerful new tools in the fight to protect them. Over the past year scientists have gathered to revise the most important international treaty aimed at conservation, the Convention on Biological Diversity (CBD). With the loss of plant and animal species accelerating, some researchers say conservation efforts should turn to remote sensing to monitor biodiversity in



Remote sensing of trees' physiological traits produced this map of functional diversity in a forest near Zürich. (Yellow is high, pink is medium, blue is low.)

near–real time across wide swaths of the globe—and help policymakers prioritize the most critical areas.

Historically, researchers had to venture out to jungles, deserts, and mountaintops to document the flora and fauna. But species distributions and abundances are changing faster than ground-based surveys can track, because of climate change, human activities, and other factors. Remote sensing offers the possibility of faster, more standardized monitoring across the entire globe. "In the past decade, there's been a revolution in the technology available to characterize ecosystems from space," says David Schimel, a research scientist at NASA's Jet Propulsion Laboratory.

Researchers are just beginning to figure out what remote sensing can and can't do and how to incorporate it into global conservation efforts. But Schimel and others see great promise for supplementing ground-based measurements with a fuller picture of ecosystems' health gleaned from aloft. "We want to transform the way biological research is done," he says.

**CAVENDER-BARES FIRST** learned that reflected or emitted light could help signal forest health during a graduate school visit to the Laboratory for the Use of Electromagnetic Radiation (LURE) at Paris-Sud University. In lab studies there almost 40 years ago, plant physiologist Jean-Marie Briantais and colleagues had shown they could measure photosynthetic activity by comparing how leaves fluoresce, or emit certain wavelengths of light, before and after being exposed to flashes of extremely bright white light. As the light displaces electrons inside chlorophyll, the intensity of red and infrared (IR) light emitted from the leaves varies, depending on how healthy the plant is. Later, Ismael Moya, a biophysicist at LURE, developed a sensor that required no extra light source, relying on sunlight instead. Moya went on to demonstrate that fluorescence could be detected from an airplane flying over crops, opening the way to surveying fields' productivity remotely. "I just became fascinated with what we could learn from the interaction of plants with light and have been for 28 years," Cavender-Bares says. In that time, researchers have mostly used methods based on reflected light, but fluorescence remains a valuable tool.

Spectral technology developed in the 1990s (bottom) advanced remote sensing and inspired ecologists such as Jeannine Cavender-Bares (top) to harness light to better understand plants and ecosystems.

PHOTOS: (TOP TO BOTTOM) MEREDITH SCHUMAN; JEANNINE CAVENDER-BARES

By now, airplanes, drones, and towers all provide spectroscopic data on vegetation. So does NASA and the U.S. Geological Survey's series of Landsat satellites, the first of which were launched in the 1970s. Initially, the agencies expected the satellites' cameras to primarily capture images in visible light, but an experimental spectral sensor on board proved the value of recording more of the electromagnetic spectrum, such as near-IR light, and by 2013 the satellites were monitoring 11 portions of the spectrum. With these "multispectral" data, researchers can monitor how "green" or productive a vegetated landscape is. Spectroscopically detected dips in chlorophyll can also signal a forest that is suffering because of drought or insect invasion—or has been cleared for development.

Cavender-Bares's occasional collaborator, remote sensing expert Michael Schaepman, aims to glean more insights from spectroscopic data. At a test area on Lägern mountain near Zürich, Schaepman—who is now president of the University of Zurich (UZH)—and colleagues have studied for decades how the beech, ash, and spruce trees there change their biochemistry throughout the seasons. In 2009, the researchers began to crisscross their study site in an airplane equipped with advanced sensors that capture 100 bands of reflected light—enough to identify the unique spectrum resulting from each tree's combination of chlorophyll, pigments, water content, and other biochemical and physical features. By cross-referencing the data with ground-based measurements of plant physiology, the researchers were able to develop computer models that turn the spectroscopic surveys into useful information about the health and diversity of the forest.

Other remote sensing methods can fill out the picture. Lidar, which typically bounces pulses from an airborne laser off the terrain below, can determine topography and the height of vegetation. (Recently, a lidar sensor was added to the International Space Station.) Based on tree heights obtained through lidar and light reflected from the plants back to a plane or satellite, Schaepman's team can calculate a forest's biomass. The researchers have also developed computer models that predict forest productivity based on changes in biomass and in the biochemical composition of the plants. And they have confirmed that satellite-borne sensors measuring long IR wavelengths can track the effects of drought. Plants normally cool their leaves by transpiring water; when they run short, their leaf temperature rises.

Schaepman hopes the lessons they've learned—and software they've developed—at Lägern can be scaled up to survey the health of larger landscapes from afar, something they are testing across 154 square kilometers of subtropical forest in China.

At the same time, he and others are also trying to zoom in—by developing remote sensing tools that capture finer detail, such as the diversity of species and communities, and even the genetics and molecular characteristics of individual trees. "My dream," says Bernhard Schmid, a UZH plant ecologist, "is that without even disturbing the system, we can measure every tree." Such fine-scale data could help fill out a picture of ecosystem health. Greater genetic diversity, for example, makes ecosystems more resilient by enabling some individuals to cope better with disease outbreaks, drought, or other environmental changes.

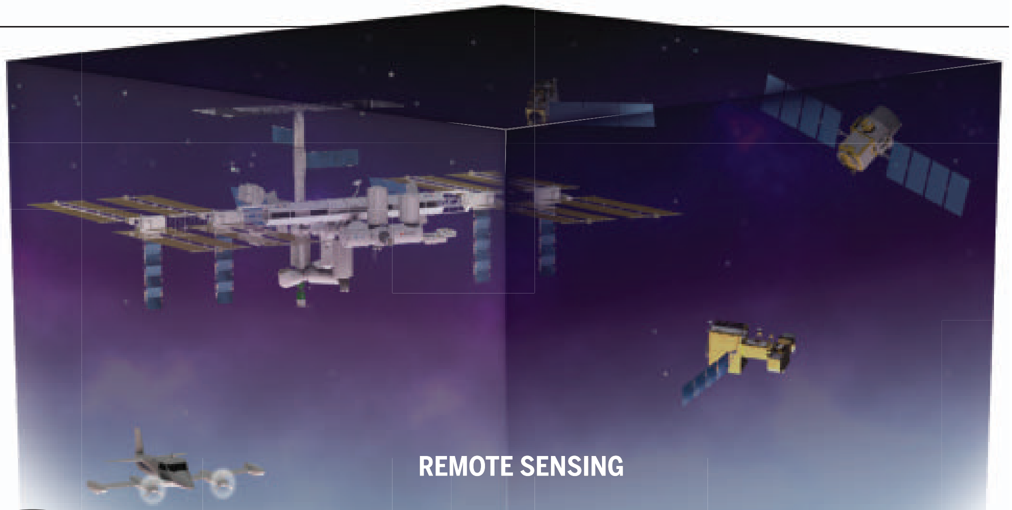
Again the clues lie in the spectrum of light reflected from foliage. Over the past decade, experiments have shown that varying concentrations of pigment, wax, and other chemicals, as well as leaf shape and other features, can subtly alter the color or brightness of the light. All of those traits have a genetic basis, and in a study published in 2020 in *Ecology and Evolution*, Schaepman's team demonstrated that spectral signatures collected from a plane matched up well with genetic data collected on the ground at Lägern.

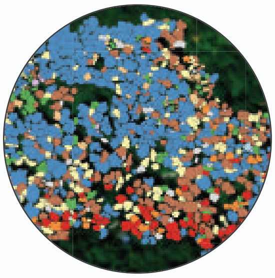
Cavender-Bares, Schaepman, and colleagues have also shown that sunlight reflected off leaves can reveal evolutionary relationships among diverse plants. Comparing spectral signatures can show which trees in a given forest are most closely related, they reported in 2020 in *New Phytologist*. The finding suggests satellite-borne sensors may one day be able to map genetic variation in a forest.

Satellite data can even uncover microbial activity and nutrient cycling in the soil—processes that also contribute to an ecosystem's overall health and productivity. At the Cedar Creek Ecosystem Science Reserve's Long-Term Ecological Research site in Minnesota, Cavender-Bares's team dug into the soil to measure respiration, microbial biomass and enzyme activity, and fungal and

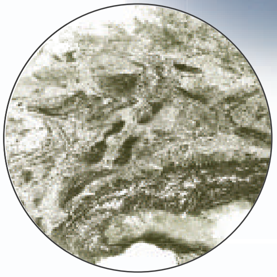
## All eyes on biodiversity

Satellites and aircraft with sophisticated sensors can track signs of ecosystem health known as essential biodiversity variables (EBVs), which are grouped into six classes. Each EBV draws on multiple types of remote sensing data, fed into computer models based on measurements from the ground. The examples below, all views of a single Swiss mountainside, show how remote sensing can address five of the six EBV classes. A sixth, genetic composition requires ground-level studies (bottom).

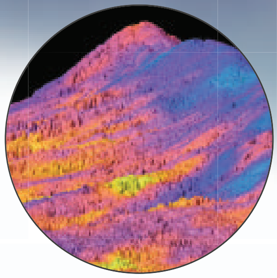




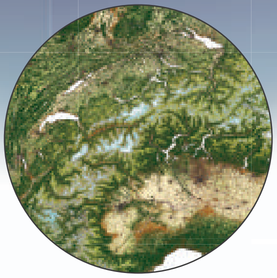
**Species by color**
Fed into models, the spectra and contours of vegetation produced this map of tree species, including beech (blue) and silver fir (red).



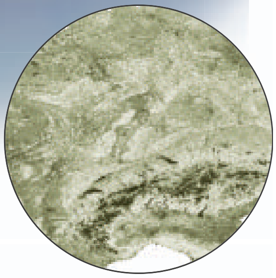
**Tree height**
Laser or radar pulses bounce back to sensors on planes or satellites, providing data on the height of the forest.



**Diversity maps**
Data on traits such as height and chlorophyll can reveal variation in a forest's functional composition.



**Land cover**
Multiple EBVs went into this map of different communities: shrublands (blue), forests (green), and cropland (beige).



**Sunlight usage**
Reflected and absorbed light provides clues about photosynthetic activity. Most active is dark, least active white.

### EBV CLASSES

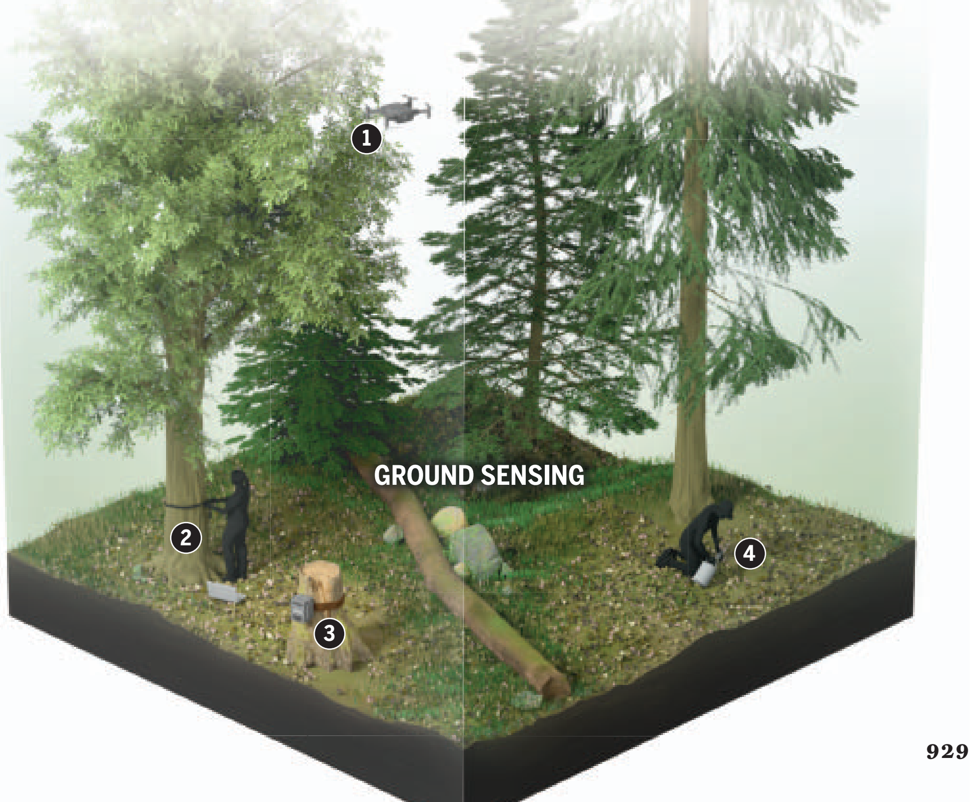
**Species population**
Remote sensing can help catalog the abundance and distribution of individual species across wide areas.

**Species traits**
The traits of some plant species, including height, physiology, and flowering time, reveal ecosystem health.

**Community composition**
A community's species diversity—a measure of resilience—varies across a region and over time.

**Ecosystem structure**
Biodiversity thrives in ecosystems with diverse land coverage, more readily monitored from above.

**Ecosystem functioning**
Photosynthetic activity sustains an ecosystem. Disturbances such as fire and disease can reduce it.



**Genetic composition**
DNA diversity is at the core of biodiversity and resilience. It is mostly determined by ground-based studies, but some new work suggests satellite data can offer clues.

**Ground sensing tools**
1 Drones for close-in remote sensing
2 Tree inventories
3 Camera trap
4 Environmental DNA sampling

CREDITS: (GRAPHIC) V. ALTOUNIAN/*SCIENCE*; (DATA) S. GRAVES *ET AL*., *REMOTE SENS*., 8, 161 (2016); MARK SIMARD/NASA-JPL; F. MORSDORF, *REMOTE SENSING OF PLANT BIODIVERSITY*, DOI.ORG/10.1007/978-3-030-33157-3_4 (2020); COPERNICUS GLOBAL LAND SERVICE

bacterial composition. The chemical makeup of decaying plants and even chemicals released by living plants into soil can affect the microbes present there, so she and her colleagues used computer modeling to look for correlations between what they found on the ground and lidar and spectral measurements of Cedar Creek's foliage. They discovered they could infer microbial activity based on remote sensing data, results they reported on 8 October in *Ecological Monographs*.

Other researchers, meanwhile, are trying to use satellite-derived information about vegetation to gain insights into the animals living there. Walter Jetz, a biodiversity scientist at Yale University, and colleagues have been developing computer models that incorporate data from the research literature about the conditions that various animal species require to thrive, such as plants used for food or nests. They feed remote sensing information about an area's climate, vegetation, and topography into the models to predict where a given animal species should be found and at what densities.

The result is the Map of Life, a freely available online repository of maps of species' ranges that draws from 600 million observations worldwide of 44,351 species of vertebrates, plants, and insects. Jetz began to build this resource 10 years ago, aiming to show how remote sensing data can help compensate for the lack of enough conservationists and ecologists to physically count every species and population throughout its range. By constantly updating the observations, Jetz's team can predict population declines or detect new threats to at-risk species.

There is even progress on eyeing animals directly from the sky. Already researchers have used satellite images to count penguins in Antarctica and follow whales and wildebeests through their migrations. And since 2018, the International Cooperation for Animal Research Using Space project has used a giant antenna attached to the International Space Station to track individual birds, mammals, and any other organisms outfitted with GPS loggers. Tags have shrunk to 4 grams and are getting smaller, meaning ever-tinier birds and even insects such as dragonflies can be followed from space.

Wearing black to minimize reflected light, researchers in Minnesota embark on a ground study that will help refine computer models for interpreting remote sensing data.

PHOTO: JEANNINE CAVENDER-BARES

**REMOTE SENSING** does have limitations. Satellite measures of biodiversity are mostly indirect, and they are limited by factors such as the spatial resolution of the satellite's sensors, how often a given area is imaged, and how many wavelengths are sensed. For biologists used to working with smaller data sets, moreover, interpreting satellite data can be a challenge. "Although there is more spectral data than ever before from airborne, satellite, and ground sensors, the data are not always easy to use," says Danielle Wood, an applied earth scientist at the Massachusetts Institute of Technology.

New satellites on the horizon, such as Landsat Next, NASA's Surface Biology and Geology mission, and the European Space Agency's (ESA's) Copernicus Hyperspectral Imaging Mission for the Environment, will have much greater capabilities than earlier Landsat satellites, capturing hundreds of bands of light. And even current technologies can be bolstered by combining different types of remote sensing, says Frank Muller-Karger, a biological oceanographer at the University of South Florida. Already, some researchers are purchasing very high-resolution imagery from private satellite companies and overlaying fine scale spectral information and lidar and radar data from other sources to reveal the forest health in ever greater detail. Both lidar and fine-scale spectral measurements from space "are really game-changing technologies, especially in combination," Jetz says.

He and other researchers stress that remote monitoring still works best if validated and enhanced with data collected on the ground, for example by community scientists, fieldwork, or monitoring efforts such as the National Ecological Observatory Network (NEON), which operates 81 sites across the United States. A team at the University of Wisconsin (UW), Madison, recently attempted to do this by combining NEON field data on birds and vegetation with remote sensing data on land cover and ecosystem health. The study, published in February in *Remote Sensing of Environment*, showed birds were most abundant where vegetation was diverse, a connection that would not have emerged from just one data set, says UW landscape ecologist and co-author Volker Radeloff.

**SUCH REMOTE SENSING** data can guide conservation efforts as well as research. In the past, governments and conservation organizations have relied primarily on expert opinion or on field studies covering only a small area to decide how much suitable habitat remains for a particular endangered species—and how much is needed. "The power of remote sensing is that it is possible to extrapolate field measurements to cover entire continents," Radeloff says.

One recent project combined remote sensing and fieldwork to map suitable habitat for the endangered Siberian white crane, *Leucogeranus leucogeranus*. The birds breed in wetlands, where the cranberries, tubers, insects, fish, and small mammals they depend on are plentiful. Field researchers working in a small 3-square-kilometer area that included some of the bird's breeding grounds had determined the cranes tend to avoid breeding where shrubs are invading the wetlands. Because the broader region is mostly inaccessible, even by foot, conservationists could only guess how much good habitat was available. "But with the satellite images, we were able to expand the area of investigation to a region of around 16,000 square kilometers,” says Claudia Röösli, a UZH remote sensing expert. With these images, conservationists can monitor the effects of floods or dry summers to assess management needs for the crane.

Many other conservation uses of remote sensing are in the proposal or pilot study phase. But wider adoption may be coming soon, if the technology is incorporated into the CBD, the landmark treaty signed in 1992 by 196 countries hoping to stem the loss of the world’s flora and fauna. “We are at a key moment,” says Nathalie Pettorelli, a conservation biologist at the Zoological Society of London. “I am hoping satellite data will gain momentum with the CBD, and I am hoping that trust in this source of information will increase.”

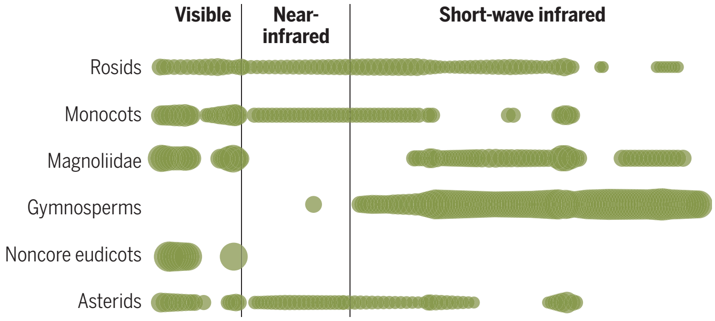
The previous CBD goals, set in 2010, included halving the rate of loss of forests and other natural habitats, slowing the spread of invasive species, and protecting coral reefs by 2020. None has been met. And the lack of a standardized way to monitor progress “was part of the reason we failed,” says Andrew Gonzalez, a conservation biologist at McGill University. That’s where some scientists think satellites and other remote sensing methods can help.

The question is exactly what should be measured and how. One potential solution follows the lead of climate scientists, who at the beginning of the 21st century established the “essential climate variables,” a set of standardized measures of carbon dioxide, temperature, and other physical characteristics that countries now use to monitor their progress toward stemming climate change. But Earth’s life is more complex than climate, says Henrique Pereira, a conservation biologist at the German Centre for Integrative Biodiversity Research, making it more challenging to document how it’s changing.

Pereira, Schaepman, and about 1500 other scientists, working under the auspices of an international organization called the Group on Earth Observations Biodiversity Observation Network (GEO BON), have spent the past 8 years developing a set of essential biodiversity variables (EBVs) for assessing and tracking changes in biodiversity across the planet. “To forge international cooperation, we need apples to apples comparisons of progress made in different countries,” says Becky Chaplin-Kramer, an ecologist at Stanford University.

## Distinct signatures

Improved sensors that detect visible, near-infrared, short-wave infrared, and other electromagnetic bands make it possible to distinguish major plant groups from space and may eventually enable remote species identification.



GRAPHIC: J.E. MEIRELES, *NEW PHYTOLOGIST*, **228**, 2 (2020), ADAPTED BY V. ALTOUNIAN/*SCIENCE*

The 20 EBVs fall into six spatial scales, from DNA to species to ecosystems (see graphic, p. 929). Remote sensing can’t monitor all of them, but it shows promise for at least the two most overarching scales, ecosystem structure and ecosystem function, Röösli says. From 2017 through 2020 she coordinated the ESA-funded GlobDiversity project, which tested the potential of satellites to provide three EBVs: the chlorophyll content of the canopy, the degree of habitat fragmentation, and the timing of the vegetation’s yearly growth cycle, known as phenology.

Röösli and her colleagues used field data from 10 reserves—including arctic tundra, wetlands, and a temperate forest—to develop and test algorithms for estimating these EBVs based on high-resolution satellite images. The results show how EBVs derived from satellite data could be used as an ecological early warning system, the researchers say. In Germany, for example, changes in the spectral signatures of spruce forests coincided with a wave of dying trees infected by an invasive spruce bark beetle.

GEO BON recommends countries or regions set up biodiversity observation networks, analogous to existing national climate observatories, that will allow each country to measure biodiversity by whatever means they have available—not just remote sensing but also ground-based methods such as camera traps and citizens counting animals. Computer models will then convert the data into EBV values and biodiversity “indicators,” which can be compared globally.

One example of such a national observatory is already being piloted in Finland. That country has embraced a mix of remote sensing and ground measurements to assess biodiversity using EBVs. Given Finland’s low population density, “remote sensing is the most promising, and perhaps only, way for comprehensive wall-to-wall ecosystem monitoring,” says Petteri Vihervaara, an ecologist at the Finnish Environment Institute.

Over the next 3 years, a Finnish Ecosystem Observatory (FEO) project will test the feasibility of determining the boundaries of wetlands, forests, and other ecosystems in Finland, in particular Lapland, using a combination of satellites, airborne lidar, aerial photographs, field surveys, and machine learning algorithms that pull all these data together to come up with an EBV called “ecosystem distribution.” Knowing where these ecosystems are is a first step toward preserving them.

In addition, FEO is using satellites to monitor water levels over time in EU-protected boreal wetlands called appa mires. This variation, an EBV called ecosystem phenology, depends on rainfall. But ditches dug nearby also affect water flow in a mire, and the remote sensing could provide an early warning of any damaging effects.

CBD negotiators are now considering proposals to incorporate indicators based on at least two EBVs into the treaty: “species distribution” to track invasive species and “ecosystem distribution” (each one’s extent and fragmentation) to monitor the richness of habitats. But whether the world is ready to embrace remote sensing in general and the EBV strategy, in particular, is unclear. Many ecologists back this approach, but it may be a harder sell for the political appointees and government representatives who need to sign off on the new CBD, because of its novelty and a lack of remote sensing resources and expertise in many lower and middle-income countries. Jetz worries that without such monitoring, countries won’t be able to track their progress toward better protection of biodiversity. But there’s not much time left to convince delegates before they meet in May 2022 to finalize the treaty.

Cavender-Bares and some other scientists aren’t convinced that reducing complex ecosystems to a small set of numbers is the best way to protect them. But regardless of whether the revised treaty embraces EBVs, Cavender-Bares is convinced remote sensing will play a critical role in tracking and combating the massive global losses now underway. “We have to be measuring biodiversity from space,” she says. ■

# INSIGHTS

EMERGENCY PATIENTS
PARKING ONLY
PARKING ONLY

POLICY FORUM

PUBLIC HEALTH

# Create a COVID-19 commission

## We need a definitive public reference for the history of events

*By* **Christopher F. Chyba**[1], **Christine K. Cassel**[2], **Susan L. Graham**[3], **John P. Holdren**[4], **Ed Penhoet**[5], **William H. Press**[6], **Maxine Savitz**[7], **Harold Varmus**[8]

The 2019 Global Health Security Index concluded that although "no country is fully prepared for epidemics or pandemics," the United States scored highest among all countries in pandemic preparedness (*1*). Yet in the COVID-19 outbreak as of October 2021, measured in deaths per 100,000 population, the United States did worse than 30 of the 35 countries defined by the International Monetary Fund as having "advanced economies," and worse than all but 18 of 184 countries ranked (*2*). Many have now called for a US COVID-19 commission to provide a definitive reference for the history of the events, to enlarge the public's understanding of them, and to prescribe strategies that might prevent, or at least mitigate, future pandemics. Drawing on our experience as members of the President's Council of Advisors on Science and Technology (PCAST) under President Obama, we present recommendations for the topics to be examined by a US COVID-19 commission (*3*, *4*).

The US Congress has created national commissions in the past to investigate events that have severely affected life, such as the 9/11 Commission. That commission gave the country a shared, bipartisan account of what had gone wrong, and why, and its recommendations led to the restructuring of the US intelligence community and the creation of the Department of Homeland Security. Similar efforts to critique national COVID-19 responses have been undertaken in other countries. For example, both the British House of Commons and the French Sénat have released detailed and critical accounts of their country's response to the pandemic (*5*, *6*).

Eight bills have been introduced in Congress so far, proposing the creation of

PHOTO: APU GOMES/AFP/GETTY IMAGES

A nurse walks into a temporary emergency room in a parking garage at Providence Cedars-Sinai Tarzana Medical Center, Tarzana, California, 3 January 2021.

such a commission (*7*). Most include only broad requirements for the issues to be investigated by the commission. Others outside of government have made similar proposals. A privately funded nongovernmental “Covid Commission Planning Group” has been formed (*8*), but has yet to offer detailed recommendations. We recognize that establishing a COVID-19 commission could be challenging in the current US domestic political environment. Nevertheless, now is an appropriate time to consider plans for the commission.

The design, composition, and resources of a national commission can determine the depth of its inquiry and the credibility of its conclusions. Congressional proposals for a COVID-19 commission have been modeled after the 9/11 Commission (*7*). Key features of the 9/11 Commission—robust budget and staffing, subpoena power, and access to previous senior officials at all levels, as well as to classified material, should be features of the COVID-19 commission.

In addition to drawing on lessons from the 9/11 Commission experience, the COVID-19 commission may wish to examine the findings of the House Select Bipartisan Committee on Hurricane Katrina, because many issues raised by the Katrina catastrophe also provide echoes of the COVID-19 experience: failed responses by the federal government despite long-standing plans, warnings, and similar recent events; failure to act on lessons from simulations; tardy exercise of federal authority; and contentious relationships between the federal and state governments.

## PUBLIC HEALTH INFRASTRUCTURE

The COVID-19 Commission will surely encounter a general issue: the undervaluing and underfunding, in the United States, of public health capabilities and practices. A prominent manifestation of this attitude has been the chronic underfunding of the Centers for Disease Control and Prevention (CDC), as well as underfunding of state and city public health agencies. The COVID-19 Commission should address how to place public health agencies in the United States on a more effective and stable footing, and it should recommend means to repair budgetary shortfalls, expand training programs for public health professions, and enhance the prestige of public health workers (*4*).

Early in the pandemic, when identification of the first infections with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was especially important, the CDC was widely criticized for its slow development and certification of a reliable test for the viral pathogen. The commission should examine those events and provide guidance about plans for developing tests for microbial pathogens to detect and monitor future outbreaks. Those tests should include methods to survey large populations for early detection of the proteins and/or nucleic acids of known and novel pathogens; to measure the presence of pathogens in asymptomatic and symptomatic people during an outbreak; and to characterize the genetic variants (like the current Delta variant) that arise and spread during continued propagation. Carrying out these tasks will require funding for the establishment and maintenance of sophisticated technical platforms, as recommended in previous reports (*3, 4*).

The adequacy of the public health informatics infrastructure and the management of public health data should be particular focuses of the commission’s attention. Public health methodology for managing data is still often manual and antiquated (*3, 4*).

## WARNING AND ORIGINS

The pace of emerging disease outbreaks such as the 2002–2004 SARS-CoV-1 pandemic, the 2009 H1N1 influenza pandemic, and the 2012 Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic had been anticipated by microbiologists and epidemiologists. In 2004, the US National Intelligence Council warned that “Some experts believe it is only a matter of time before a new pandemic appears [that] would be devastating and could spread rapidly throughout the world” (*9*). A 2008 study cataloged the emergence of 335 infectious diseases in humans between 1940 and 2004 (*10*). Nearly all subsided quickly. But it was clear that the threat was ever-present, and likely increasing.

A COVID-19 commission should review those earlier outbreaks, their causes, and the US responses and determine whether there is a “panic/neglect cycle” of increased preparedness after each episode that then lapses until the next outbreak induces another burst of attention. If so, the commission should propose mechanisms to ensure more consistent attention to such threats.

It is important to understand and report on the roles played by the scientific community, other governments, and international organizations at the early stages of the COVID-19 pandemic. When and how did the United States receive warning from China of the initial cases in Wuhan and of the threat they might pose? How did this information move through the US government, and how and when did it reach the White House and the president? Since 2017, reportedly more than 30 staff members had been pulled out of the CDC’s office in China, including a US epidemiologist embedded in China’s equivalent to the CDC (*11*). If they had still been in place, could these personnel have provided earlier warnings about

---

[1]Department of Astrophysical Sciences and School of Public and International Affairs, Princeton University, Princeton, NJ, USA. [2]Department of Medicine, School of Medicine, University of California, San Francisco, CA, USA.[3]Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, USA.[4]Kennedy School of Government and Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA.[5]Division of Biological Sciences, University of California, Berkeley, CA, USA. [6]Department of Computer Science and Integrative Biology, The University of Texas, Austin TX, USA.[7]National Academy of Engineering, Washington, DC, USA. [8]Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10021, USA. Email: cchyba@princeton.edu

the COVID outbreak in Wuhan and its character? Was there appropriate communication among US health, diplomatic, and other agency personnel, and with the World Health Organization (WHO)?

It will be important to decide whether the commission's mandate includes a review of public and classified information about the origin of the pandemic. This topic risks dominating the attention paid to the commission's final report, potentially undermining other important findings. At the same time, if there were policy conclusions and recommendations to be made regarding gain-of-function experiments or high-containment laboratory practices, these could be of importance and standing regardless of the resolution of the COVID-19 origins issue.

## PREPAREDNESS

The commission should review the adequacy of pandemic planning in the United States prior to the COVID-19 outbreak, including the adequacy of the nation's strategic stockpile inventory and its governing strategy, the supply chain's readiness to meet manufacturing standards and demands, and the availability of research and clinical trial methodologies to respond to a new infectious agent by developing, testing, and producing tests, therapeutics, and vaccines in a coordinated and rapid fashion (*4*).

In 2019, the US government, in cooperation with state, local, and tribal health authorities, conducted an outbreak simulation exercise. The after-action report from the exercise presciently highlighted many challenges that emerged in practice in the COVID-19 case; it has now been released under a Freedom of Information Act request (*12*). The commission should examine lessons learned from the exercise, evaluate the impact of its findings within the US government, and recommend how to make it more likely that findings from future exercises are acted upon.

It has long been argued that preparations for bioterrorism or biological attack by a state adversary should have many synergies with preparedness for naturally occurring disease, a point reemphasized by PCAST in its biodefense report to President Obama (*3*). The commission should consider the proper balance, synergies, and potential risk trade-offs among strategies for protection against natural occurring disease, laboratory accident, and biological attack. The commission should also ask whether the US government had this balance right at the time of the COVID-19 outbreak.

Actions taken by the federal government well before the appearance of COVID-19 warrant close examination. For example, in May 2018, the Directorate for Global Health Security and Biodefense within the National Security Council staff, which had been created in 2015, was disbanded (*4*). How consequential was this reorganization with respect to US ability to anticipate and respond to the COVID pandemic? Similarly, in 2017, the incoming administration reportedly was provided with the National Security Council's "Playbook for Early Response to High-Consequence Emerging Infectious Disease Threats and Biological Incidents" (*13*). To what extent were key members in the new administration aware of the document, and to what extent were its prescriptions followed? To the extent they were not, was that because they were judged to be flawed, or for other reasons?

A: FULL
B: DNR
C: FULL
D: FULL
RT:
23470
Mariner
Fent
ZOD
PRONING TEAM
310.294.6473
NEXT PRONE: 1/21 1700
NEXT SUPINE:

A clinician cares for COVID-19 patients in a makeshift Intensive Care Unit at Harbor-UCLA Medical Center, Torrance, California, 31 January 2021.

PHOTO: MARIO TAMA/GETTY IMAGES

Other events affected the capacity of the United States to be prepared to provide needed supplies for patient care and personal protection during the COVID-19 pandemic. The Strategic National Stockpile (SNS) was depleted during the H1N1 epidemic in 2009. In March 2013, Congress reauthorized the SNS for financial year (FY) 2014–FY 2018 (*4*). Why did this authorization not lead to appropriations? Why was the authorization not renewed after 2018? SNS planning assumed that state stockpiles would be in place. For the most part, however, those stockpiles were also not replenished and maintained after the 2008 financial crisis (*4*). This failure should also be understood, with lessons drawn for future readiness. Could greater attention to supply-chain issues have helped to replenish the stockpile? Was the United States too dependent on foreign manufacturers for items in its stockpile, and, if so, how could this vulnerability have been mitigated—for instance, by better use of the Defense Production Act?

A major component of the Commission's attention to the question of preparedness should be an examination of the extent to which the CDC was unprepared for an epidemic of COVID-19's scale, and if so, why.

## RESPONSE TO THE ARRIVAL OF THE PANDEMIC IN THE UNITED STATES

A COVID-19 commission will need to evaluate the ways in which various components of the nation's medical infrastructure—public health agencies, health care facilities, public and private research institutions, and regulatory bodies—responded to the emergency.

The commission should examine the decision-making process that led to restrictions prohibiting travelers from entering the United States and should assess their public health effectiveness. Should there have been all-inclusive bans, including indirect travel through other countries? Or should there have been no bans? Should there have been mandatory quarantine? The commission should also examine the efficacy of domestic actions that were taken by individual states within the United States, such as travel advisories and recommended quarantine after arrival from certain states.

In the absence of effective vaccines or therapies to curtail the spread and morbidity of a novel viral pathogen, all nations were obliged to depend on traditional means to combat the pandemic by limiting the spread of the pathogen, such as social distancing and mask wearing. Each of these strategies and the means to impose it has been debated widely and will be important subjects for the commission's study. The commission should also examine the nation's capacity to provide the funds and trained personnel to perform contact tracing, whether there were deficiencies that had a notable role in the failure to control the pandemic, and whether resources were inappropriately expended on contact tracing after high case rates had rendered it infeasible (*4*).

The commission should examine US government efforts to quickly establish a vaccine research initiative and provide financial incentives to private industries, in hopes of accelerating the development, testing, and production of a vaccine. The inquiry should identify the previous research underlying capacity to develop, produce, and distribute vaccines, together with production and distribution efforts (nationally and globally), and draw lessons both for future outbreaks and for vaccine production under noncrisis conditions.

Public health measures failed to prevent widespread transmission of SARS-CoV-2 in the United States. At its peaks, this situation placed enormous stresses on US hospitals, other health care delivery sites, frontline personnel, and provision of materials. The commission should examine how well the challenge of caring for so many patients seriously affected by COVID-19 was met, how the pandemic affected the fiscal status of hospitals, and how the nation's health care facilities can be better prepared for future pandemics. It should also address what information should be offered to patients and families in times and places where standard-of-care is compromised because of overload.

The therapeutic arsenal that might have reduced the severity of COVID-19 and its mortality rates remains small, despite extensive efforts to identify existing US Food and Drug Administration (FDA)–approved drugs, to find new ones, or to develop immunological therapies that are beneficial and safe. The commission will need to review these efforts and examine whether the FDA's approval processes were properly free of political influence.

One unexpected aspect of COVID-19 has been the complexity of its pathogenic process. The commission should evaluate the efforts made by the biomedical research community to evaluate the mechanisms of pathogenesis and to identify both genetic and environmental factors that contribute to the severity and longevity of the illness.

The role and effects of the US and state governments, news programs and websites, scientists and doctors, and social media in spreading information, misinformation, and unsupported claims regarding the pandemic should be examined and illuminated. The commission should produce a classified annex on the role of any foreign nations' misinformation campaigns targeted at government leaders, news programs, or the public.

## HEALTH EQUITY ISSUES

The outcome of infection is now known to be adversely affected by advanced age, obesity, and immune systems already compromised by cancer and other afflictions, among other patient characteristics. Populations with a high prevalence of these conditions are at special risk. For example, nearly one-third of COVID-19 deaths in the United States through June 2021 were linked to nursing homes (*14*). The commission should examine how these institutions could better have protected their residents and staff. Similar questions should be asked about front-line workers in hospitals, essential workers in other professions, and prison staff and inmates.

An especially high burden of disease has been recorded among some ethnic minorities. Black, Hispanic, and Indigenous Americans have been twice as likely to die from COVID-19 as white or Asian Americans (*15*). The commission should probe the nature of and reasons for these disparities, investigate the extent to which they result from broader health inequity issues, examine whether the United States collects the type of public health and social demographic data needed to adequately address these issues, and make recommendations aimed at reducing these disparities.

## CONCLUSION

The task of a nonpartisan COVID-19 commission will be to produce a clear-eyed assessment of why and how the United States fared so poorly in this pandemic, as well as how particular successes were achieved. Where possible, it will be valuable to make this assessment in a comparative context, both with respect to previous pandemics and with respect to the experiences of other advanced-economy nations (*2*, *5*, *6*). The commission should identify how the relevant US institutions can be improved (potentially by both legislative and executive action) and better work together to ensure superior outcomes when future pandemics occur, as they surely will. ■

### REFERENCES AND NOTES

1. Nuclear Threat Initiative and Center for Health Security, Johns Hopkins Bloomberg School of Public Health, Global Health Security Index (2019); www.ghsindex.org/wp-content/uploads/2020/04/2019-Global-Health-Security-Index.pdf.
2. Johns Hopkins University and Medicine Coronavirus Resource Center, "Mortality Analysis" (23 October 2021); https://coronavirus.jhu.edu/data/mortality.
3. PCAST Document and Reports; https://obamawhitehouse.archives.gov/administration/eop/ostp/pcast/docsreports/archives.
4. OPCAST Ad Hoc Pandemic Response Group, https://opcast.org.
5. Health and Social Care, and Science and Technology Committees (12 October 2021), *Coronavirus: Lessons learned to date. Sixth Report of the Health and Social Care Committee and Third Report of the Science and Technology Committee of Session 2021–22* (2021-2022 HC 92), House of Commons, London; https://committees.parliament.uk/work/657/coronavirus-lessons-learnt/.
6. Sénat (8 December 2020), *Rapport Fait au nom de la commission d'enquête (1) pour l'évaluation des politiques publiques face aux grandes pandémies à la lumière de la crise sanitaire de la covid-19 et de sa gestion* (Session ordinaire de 2020-2021 N° 199), Parlement français, Paris; www.senat.fr/fileadmin/Fichiers/Images/redaction_multimedia/2020/2020-Documents_pdf/20201210-Conf_presse_Covid_19/20201210_rapport_CE_Covid__tome_1__conf-presse.pdf.
7. Congressional Research Service, Proposals for a COVID-19 Congressional Advisory Commission in the 116th Congress: A Comparative Analysis (28 January 2021); https://crsreports.congress.gov/product/pdf/R/R46330.
8. Miller Center, University of Virginia, *Covid Commission Planning Group*; https://millercenter.org/covidcpg.
9. National Intelligence Council, Mapping the Global Future: Report of the National Intelligence Council's 2020 Project, NIC 2004-13 (December 2004), p. 30; www.dni.gov/files/documents/Global%20Trends_Mapping%20the%20Global%20Future%202020%20Project.pdf.
10. K. E. Jones *et al.*, *Nature* **451**, 990 (2008).
11. M. Taylor, "Exclusive: U.S. Slashed CDC Staff inside China Prior to Coronavirus Outbreak," *Reuters*, 26 March 2020; www.reuters.com/article/us-health-coronavirus-china-cdc-exclusiv-idUSKBN21C3N5.
12. U.S. Department of Health and Human Services, *Crimson Contagion 2019 Functional Exercise After-Action Report* (January 2020); www.governmentattic.org/38docs/HHSaarCrimsonContAAR_2020.pdf.
13. D. Diamond, N. Toosi, "Trump team failed to follow NSC's pandemic playbook", *Politico,* 25 March 2020*;* www.politico.com/news/2020/03/25/trump-coronavirus-national-security-council-149285 (2020).
14. The New York Times, "Nearly One-Third of U.S. Coronavirus Deaths are Linked to Nursing Homes," *The New York Times*, 1 June 2021; www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html.
15. APM Research Lab, "The Color of Coronavirus: COVID-19 Deaths by Race and Ethnicity in the U.S.," 5 March 2021; www.apmresearchlab.org/covid/deaths-by-race.

### ACKNOWLEDGMENTS

We are grateful for conversations with R. Levin. C.F.C. thanks E. Magistro and G. Z. McDermott for research assistance. C.F.C. was primary author for the initial and revised manuscripts. All authors contributed to writing and editing each manuscript. E.P. is director of the UCSF Benioff Children's Hospitals and has a financial interest in Alta Partners, a venture capital firm with investments in several biotechnology companies. All other authors declare no competing interests.

10.1126/science.abk0029

PERSPECTIVES

PHYSICS

# No vacancy in the Fermi sea

## Optical transparency is increased by the Pauli principle in ultracold atom gases

By **Brian DeMarco**[1] and **Joseph H. Thywissen**[2]

Light scattering is one of the most effective scientific probes of matter. When light is used to investigate materials, the influence on a source, such as a laser beam, is measured. On pages 979, 972, and 976 of this issue, Sanner *et al.* (*1*), Deb and Kjaergaard (*2*), and Margalit *et al.* (*3*), respectively, report on the observation of a previously unobserved way in which the quantum exchange symmetry of neutral atoms is imprinted on the intensity and distribution of scattered photons.

The interaction of light with matter gives rise to a wide range of phenomena. For example, light can be scattered elastically from one direction into another. This effect is used in x-ray crystallography, which was used to identify the structure of DNA. Light can also scatter inelastically and change its wavelength through the Raman effect, which provides a fingerprint of the chemical composition of materials. Quantum states generate an even richer tapestry of scattering behavior that provides insights into the behavior of subatomic particles and electrons in materials.

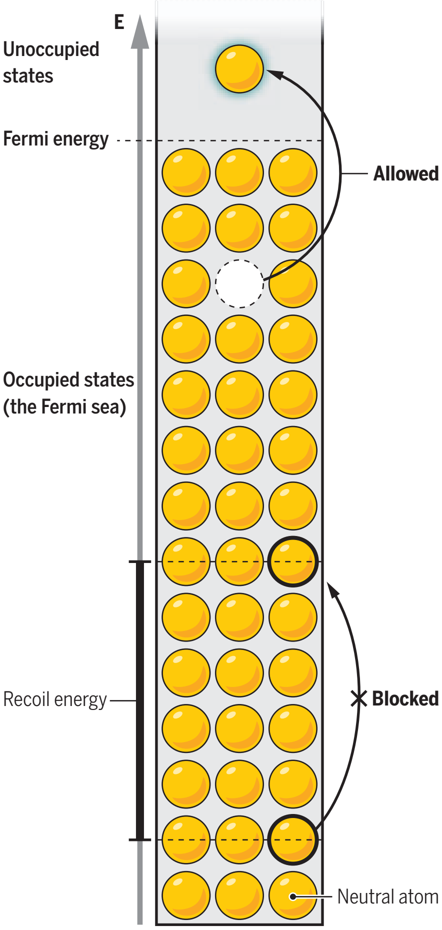
A previously unobserved quantum effect has been measured in these experiments—the Pauli exclusion principle suppressing light scattering in a gas of fermionic atoms cooled to ultralow temperatures. The Pauli principle is familiar as the rule that underlies the periodic table—no two bound electrons can occupy the same quantum state. This law is a consequence of exchange symmetry, which constrains the states of the two types of fundamental quantum particles found in the universe, bosons and fermions.

The exclusion principle applies to all collections of indistinguishable fermionic particles, including trapped gases of ultracold potassium, lithium, or strontium atoms. These atoms are composite fermions because the quantum number describing the combined spin of their neutrons, protons, and electrons is a half-integer. As in metals and neutron stars, when atomic gases are sufficiently cooled, a "Fermi sea" emerges. At absolute zero, all quantum states with an energy below the Fermi energy are occupied (see the figure).

The development of this collective quantum state changes the material properties. For neutron stars, this effect is drastic, and Fermi degeneracy stabilizes against gravitational collapse through quantum pressure. The impact is more subtle in ultracold atomic gases. One manifestation is the suppression of light scattering, which was proposed as a method to detect Fermi degeneracy more than 20 years ago. Light scattering is reduced in this regime because it involves recoil, or transfer of momentum and energy from the light to the atom. When the Fermi sea is full, this transfer cannot occur because the final states of the atoms are already occupied, except for a small band around the Fermi energy. The Pauli principle therefore forbids many light-scattering events.

### Squashing light scattering

Light scattering can promote neutral atoms to a higher energy level. In a system of fermions, each quantum state can hold only one atom. This means that lower-energy atoms may not have an empty state to be excited into because all accessible states are already occupied. The net effect of this is to decrease the amount of light scattering.



GRAPHIC: A. MASTIN/*SCIENCE*

Three research teams have now reported the observation of this effect. All three teams observed that the rate of light scattering was reduced when a trapped atomic cloud was cooled to ultracold temperatures.

Sanner *et al.*, at the JILA in Boulder, Colorado, measured the rate of light scattering from fermionic strontium ($^{87}$Sr). The authors observed that when a Fermi sea is formed, forward-scattered light is suppressed far more than side-scattered light. This behavior arises because atoms recoil more if they are kicked by larger-angle Rayleigh scattering and are therefore more easily able to find an empty final state outside the Fermi sea.

Deb and Kjaergaard at the University of Otago measured the extinction of light transmitted through fermionic potassium ($^{40}$K). The blocking effect manifests as an increased transparency of the cloud. By comparison, the transparency of a bosonic cloud of rubidium is not increased at low temperature. The exchange symmetry of bosons allows for multiple occupation of the same quantum state, so the Pauli principle should not apply. The authors also observe that the phase shift of a laser beam probe crossing the $^{40}$K gas shows no signature of blocking, demonstrating that coherent processes are not Pauli suppressed.

Margalit *et al.*, at the Massachusetts Institute of Technology, studied a dense cloud of fermionic lithium ($^{6}$Li). The higher ratio of Fermi energy to photon recoil energy enabled them to observe stronger

---

[1] Department of Physics and Illinois Quantum Information Science and Technology Center (IQUIST), University of Illinois at Urbana-Champaign, Urbana, IL, USA.
[2] Department of Physics, University of Toronto, Toronto, ON M5S1A7, Canada. Email: bdemarco@illinois.edu; joseph.thywissen@utoronto.ca

suppression of photons scattered at large angles. The authors also demonstrate the fragility of an ultracold Fermi sea, as only one photon can be detected per thousand atoms within the blocking regime. Beyond this, the phenomenon starts to fade because heating increases the number of unoccupied states, thereby increasing the available final states.

Margalit *et al.* present an alternative explanation for their observations, which can be applied to all three experimental results. Light scattering is the result of density fluctuations. For example, the white color of milk is the result of its heterogeneous density at the wavelength scale, created by a suspension of micelles. Similarly, the suppression of light scattering in these ultracold gases is the result of the sub-Poissonian density fluctuation because fermionic statistics suppress the probability that two fermions approach each other closer than their de Broglie wavelength. This "smooths out" the density of the gas. Taken to an extreme, a perfectly homogeneous material would be fully transparent.

All of the authors avoided resonant fluorescence by working with off-resonant laser light. This means that the laser wavelength was tuned far enough away from the atomic resonance so that the probability of scattering per photon was small. Using this strategy avoids the complications of multiple scattering, which is complex and has yielded its own recent surprises, such as the prediction of a maximum index of refraction through renormalization group theory ideas (*4*). However, the phenomenon as originally discussed (*5*), suppression of spontaneous emission, is yet to be observed.

The observation of a phenomenon that involves the collective quantum state of the Fermi sea is an important milestone for quantum-gas physics. Experiments can now leverage this effect as a tool to help cool, manipulate, and measure the properties of ultracold atoms in different ways. There is also the potential for using Pauli suppression of light scattering to improve the performance of atom-based quantum devices, such as optical lattice clocks and quantum network nodes. ■

**REFERENCES AND NOTES**

1. C. Sanner, L. Sonderhouse, R. B. Hutson, L. Yan, W. R. Milner, J. Ye, *Science* **374**, 979 (2021).
2. A. B. Deb, N. Kjærgaard, *Science* **374**, 972 (2021).
3. Y. Margalit, Y.-K. Lu, F. Ç. Top, W. Ketterle, *Science* **374**, 976 (2021).
4. F. Andreoli, M. J. Gullans, A. A. High, A. Browaeys, D. E. Chan, *Physical Review X* **11**, 011026 (2021).
5. T. Busch, J. Anglin, I. Cirac, P. Zoller, *Europhysical Letters* **44**, 1 (1998).

10.1126/science.abm0072

NEUROSCIENCE

# The head and the heart of fear

## The insular cortex calibrates brain, body, and behavioral responses to danger

*By* **John P. Christianson**

Have you ever caught yourself speaking too loudly and quickly begin to whisper? This adjustment requires you to connect the volume of your voice, through brain commands to the vocal cords and diaphragm, with feedback through your own ears. Refinements like this occur for many behaviors and afford adaptation to a wide range of situations. On page 1010 of this issue, Klein *et al.* (*1*) investigated how and where this feedback integration occurs in response to threat. Akin to modulating our voice to be heard clearly, adjusting behavioral responses to environmental danger involves feedback integration. Insufficient reactions to danger may lead to risky behaviors, whereas overly strong reactions at times of safety would prevent exploration, feeding, or mating. The latter extreme is manifest in posttraumatic stress disorder, where feelings of dread and avoidance behaviors often prevail despite evidence of safety (*2*).

When expecting danger, animals display a combination of autonomic (e.g., changes in heart rate, breathing), behavioral (e.g., fight or flight), and cognitive (e.g., feelings of dread) responses for self-preservation. These are interrelated at the biological level such that a change in heart rate can modify behavioral and cognitive responses to danger, and vice versa. The emotional experience of fear emerges from the moment-to-moment interplay among mental appraisal and sensations from external and internal danger cues (*3*). Thus, the behavioral and mental aspects of emotion are subject to constant updating that is informed by incoming information and cognitive perceptions. The study of Klein *et al.* exposes some of the biological machinery through which the autonomic, behavioral, and mental aspects of the threat response are integrated (see the figure).

Laboratory mice were taught, or conditioned to, a danger signal through repeated presentation of a sound with a mild footshock. Subsequent presentations of the danger signal will evoke mental predictions that threat is imminent and evoke changes in heart rate and behavioral freezing. However, further presentation of the danger cue without shock violates the prediction of danger and leads to a normalization, or extinction, of cue-elicited autonomic and behavioral responses.

Searching for the neural loci wherein internal bodily states—such as heart rate, behavioral outputs (freezing), and cognition—merge, Klein *et al.* focused on the insular cortex. The insula is anatomically positioned to process visceral information (*4*), is interconnected with cortical and subcortical networks involved in executive function and behavioral control (*5*), and is implicated in the learning about danger and safety (*6*). Using optogenetics, a technology that provides temporally precise control of neural activity, the authors inhibited insula neurons during danger signals under conditions that led to either weak or strong danger learning. In mice exhibiting appropriately low freezing behavior after weak danger learning, insula inhibition increased freezing. However, in mice exhibiting appropriately high freezing after strong danger learning, insula inhibition reduced freezing. The interpretation is that the insular cortex contributes to the fine-tuning of behavioral responses to match the level of expected danger.

Cognitive and behavioral responses to danger cues are influenced by internal states, including heart rate. Central regulation and sensitivity to heartbeat is mediated by atrial baroreceptors (blood pressure sensors) that convey signals through the vagus nerve to the brainstem and, ultimately, to the insular cortex (*7*). Klein *et al.* hypothesized that in response to danger signals, information about heart rate and behavioral freezing might be integrated in the insula where it could be used to update or adjust danger responses. A key innovation in this study involved simultaneous recording of heart rate and insular cortex neuronal activity during danger learning and extinction. As with behavioral freezing and heart rate, insula neurons responding to the danger cue increased with sound-shock pairings, then reduced during extinction. Notably, the activity of insular neurons in response to the danger cue correlated closely with variations in heart rate. Synthetic stimulation of the vagus nerve to scramble in-

---

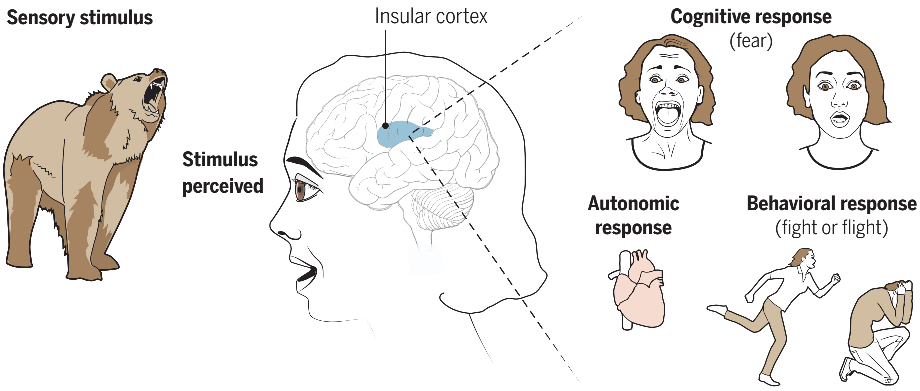
Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA, USA. Email: j.christianson@bc.edu

coming heart rate information selectively affected the freezing behavior and insular activity of mice at either the high or the low extremes of freezing behavior, indicating that vagal input is necessary for calibrating danger responses.

These new findings add to growing knowledge about the insula as a site of convergence for autonomic, somatic, and cognitive information to shape decision-making. In turn, insula outputs shape consumption, approach or avoidance, and defensive behaviors so that they are adaptive for a given circumstance. For instance, when hungry, the insula enhances the salience of food-predictive stimuli (*8*), and the insula contributes to approach or avoidance decisions when surrounded by stressed individuals (*9*). During extinction, mental predictions about the likelihood of danger must be updated, which requires the insula (*1*). These examples of insular contribution to behavior involve ambiguity or uncertainty, suggesting that the insular cortex is at the apex of a process where predictions are made about bodily states, with its output tuning subcortical processes (*10*). Alternatively, the insula may be a link in a cortical decision-making circuit that influences behavior indirectly by recruiting additional executive resources to resolve ambiguous situations. Existing data support both ideas—artificial stimulation of insula output has been found to be sufficient to alter social approach, freezing, and consummatory behavior (*11*, *12*), and functional neuroimaging studies have found the insula to be central to a "salience" network that responds to ambiguity to bias central nervous system activity toward executive function (*5*).

### Feedback integration in response to threat

When expecting danger, animals display a combination of autonomic (e.g., changes in heart rate, breathing), behavioral (e.g., fight or flight), and cognitive (e.g., feelings of dread) responses for self-preservation. Klein *et al*. find that these features merge on the insular cortex in mice, which fine-tunes behavioral responses to match the level of expected danger.



GRAPHIC: N. CARY/SCIENCE

Turning away from the laboratory to the increasingly "virtual" nature of society, the need to understand the mental, behavioral, and bodily aspects of emotion is as important as ever. The system under study in the mouse evolved to navigate both overt and perceived dangers in nature. Humans must use this same system where physical harm is relatively unlikely but psychological harms in the form of social, economic, and other stressors are increasingly identified as drivers of anxiety and mental illness. Encouragingly, mental awareness of internal states, also referred to as "interoceptive ability," can be learned (*4*) and may help individuals better regulate emotions and increase stress resilience (*13*). The work of Klein *et al.* underscores the importance of a holistic, "mind, brain, and body" approach to understanding emotion and provides a template for future basic science to uncover how complex states associated with pain and sickness interact with mental processes. ■

**REFERENCES AND NOTES**

1. A. S. Klein, N. N. Dolensek, C. Weiand, N. Gogolla, *Science* **374**, 1010 (2021).
2. S. Rauch, L. M. Shin, E. A. Phelps, *Biol. Psychiatry* **60**, 376 (2006).
3. L. Feldman Barrett *et al.*, *Annu. Rev. Psychol*. **58**, 373 (2007).
4. S. S. Khalsa *et al.*, *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 501 (2018).
5. L. Q. Uddin, *Nat. Rev. Neurosci*. **16**, 55 (2015).
6. A. R. Foilb, J. G. Flyer-Adams, S. F. Maier, J. P. Christianson, *Neurobiol. Learn. Mem*. **134**, 317 (2016).
7. S. N. Garfinkel, H. D. Critchley, *Trends Cogn. Sci*. **20**, 34 (2016).
8. Y. Livneh *et al.*, *Nature* **546**, 611 (2017).
9. M. M. Rogers-Carter *et al.*, *Nat. Neurosci*. **21**, 404 (2018).
10. Y. Livneh, M. L. Andermann, *Neuron* 10.1016/j.neuron.2021.08.036 (2021).
11. D. Gehrlach *et al.*, *Nat. Neurosci*. **22**, 1424 (2019).
12. M. R. Rogers-Carter *et al.*, *J. Neurosci*. **7**, 316 (2019).
13. L. Haase *et al.*, *Biol. Psychol*. **113**, 37 (2016).

10.1126/science.abm6790

---

QUANTUM MECHANICS

# A molecular double-slit experiment

## Quantum interaction similar to light interference is observed in molecular collisions

*By* **Xingan Wang**[1,2] *and* **Xueming Yang**[3,4,5]

Isaac Newton and Christiaan Huygens debated whether light is a wave or a particle. The disagreement was seemingly settled in 1801 when Thomas Young demonstrated the wave nature of light in the double-slit experiment (*1*). This false dichotomous understanding of light would continue until 1927, when Clinton Davisson and Lester Germer showed the possibility of wave-particle duality through electrons, ushering in perhaps the most important discovery in quantum mechanics (*2*). Wave-particle duality has since been demonstrated in countless experiments using different systems. However, almost a century later, scientists are still grappling with how to predict the effects of wave-like interference in more sophisticated systems, such as those involving the nonbinding interactions between atoms and molecules known as steric effects. On page 960 of this issue, Zhou *et al.* (*3*) present observations of a low-energy inelastic bimolecular collision between sterically controlled deuterium molecules ($D_2$) and helium atoms (He).

Using an experimental setup that enabled the investigation to be performed under single-collision conditions, Zhou *et al.* measured the inelastic rotational energy transfer process between $D_2$ and He. For half of the experiment, $D_2$ molecules in the biaxial state were collided with He atoms,

---

[1]Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei 230026, China. [2]Department of Chemical Physics, University of Science and Technology of China, Hefei 230026, China. [3]Department of Chemistry, Southern University of Science and Technology, Shenzhen 518055, China. [4]Center for Advanced Light Source, Southern University of Science and Technology, Shenzhen 518055, China. [5]State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China. Email: xawang@ustc.edu.cn; yangxm@sustech.edu.cn

whereas for the other half, D$_2$ molecules in the uniaxial state were used. In both cases, the scattered D$_2$ molecules were detected using resonance-enhanced multiphoton ionization, and their velocity distribution was determined by time-of-flight mass spectrometry. In addition, because the center-of-mass frame was coincident with the moving frame of reference of the incoming D$_2$ molecules as they collide with the He atoms, the authors could extract the scattering angular distribution of the scattered D$_2$ molecules from the experimentally measured velocity distribution.

When preparing the D$_2$ molecules for the uniaxial states, the bond axes of the molecules were tilted to +45° or −45° from the center-of-mass collision axis. When preparing the molecules for the biaxial state, however, the bond axis of the biaxial state molecules becomes a superposition of both +45° and −45° (see the figure).

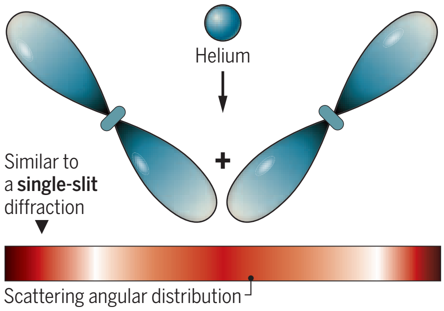
In quantum mechanical terms, the scattering of the biaxial state should have two indistinguishable collision pathways that interact with each other through quantum interference. Zhou *et al.* observed a clear difference between the inelastic scattering angular distribution of the biaxial state compared with the two uniaxial states. In the case of the two uniaxial states, the scattering between D$_2$ and He can only occur with one uniaxial D$_2$ molecule at a time. This produces a scattering angular distribution similar to a single-slit diffraction pattern using light. Conversely, the biaxial state D$_2$ is a superposition state that contains both orientations and thus produces a scattering angular distribution similar to a double-slit interference pattern for light. This direct comparison of the distinct scattering pattern demonstrates strong quantum interference between the biaxial D$_2$ molecules and the He atoms during their collision. In all, the experiment provides an elegant realization of quantum interference in molecular collision processes, analogous to the famous double-slit experiment for light performed by Thomas Young more than two centuries ago (*1*).

## Double-slit interference pattern observed in molecular collision

Deuterium molecules (D$_2$) are set in the uniaxial state and the biaxial state before being sent on their way to collide with helium atoms. The biaxial D$_2$, represented by the superposition of the two uniaxial states, shows a noticeably different scattering pattern from the uniaxial D$_2$, thus demonstrating the presence of strong quantum interference in the biaxial D$_2$ during collision with helium.

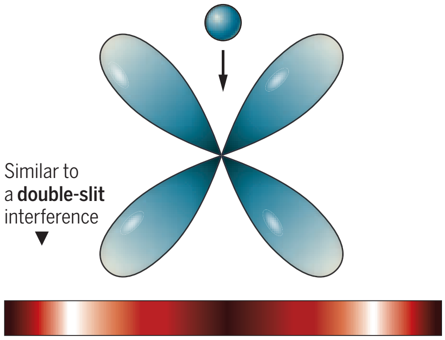
### Two uniaxial states at +45° and −45°

D$_2$ bond axes were tilted from the center-of-mass collision axis


Helium
+
Similar to a single-slit diffraction
Scattering angular distribution

### A single biaxial state at ±45°

A superposition state that contains both orientations



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Zhou *et al.* are not alone in the hunt for double-slit analogs in nonoptical systems. In recent years, various quantum interference phenomena in inelastic and reactive collisions have been reported, such as the observation of quantum interference patterns in nitric oxide molecules scattering with noble gas atoms (*4*, *5*). Quantum interference has also been reported in the forward scattering direction of the H + HD → H$_2$ + D reactive collision, with the relationship between quantum interference of quantized transition states analogous to the slits in a double-slit experiment (*6*–*8*). Zhou *et al.* add another member to the roster of double-slit analogs and provide a distinctive platform for investigating stereodynamics in bimolecular collisions.

Besides the originality of observing quantum interference in an exotic system, the study of the quantum dynamics in atomic and molecular collisions with aligned or oriented molecules also advances the understanding of molecular energy transfer and reaction dynamics. One obvious choice for further study in this field is triatomic hydrogen (H$_3$), which offers an intrinsic quantum double-slit because of its conical intersection. Among the processes involving H$_3$ systems, the H + HD → H$_2$ + D reaction has been observed to exhibit quantum interference at a collision energy near the well-known conical intersection during scattering (*9*). The presence of a conical intersection between two adjacent electronic states in the potential energy surface creates two distinctive topological reaction pathways. The two reaction pathways, clockwise and counterclockwise, around the conical intersection could act as a double slit in the chemical reaction, with one corresponding to the direct reaction and the other corresponding to a roaming-like pathway. Interferences between the two distinctive topological reaction pathways play a key role in unraveling the geometric phase effect in this benchmark chemical reaction (*10*).

The experimental setup used by Zhou *et al.* provides an efficient method to prepare spatially aligned H$_2$, HD, and D$_2$ molecules. During their experiment, the D$_2$ molecules in the rotational and vibrational ground states were brought to the excited state with near 100% efficiency. (The efficiency was enabled by using the Stark-induced adiabatic Raman passage facilitated by a pump laser and a Stokes laser with linear polarizations.) In their setup, Zhou *et al.* demonstrated an effective way to control the alignment of the D$_2$ molecules through different combinations of the pump laser polarization and the Stokes laser polarization. This technique can be useful in future experiments that seek to investigate the stereodynamics of chemical reactions, that is, how the alignment or orientation of chemical bonds affects molecular collisions and chemical reactions. For example, future scattering experiments that use crossed molecular beams may provide key insights on the dynamical origin of steric effects in chemical processes such as the H + HD → H$_2$ + D reaction. The "double-slit" approach provides a new way to coherently control molecular collisions through the alignment or orientation of bond axes and may lead to other opportunities in the quantum control of chemical reactions. ■

### REFERENCES AND NOTES

1. T. Young, *Philos. Trans. R. Soc. London* **94**, 1 (1804).
2. C. J. Davisson, L. H. Germer, *Phys. Rev*. **30**, 705 (1927).
3. H. Zhou, W. E. Perreault, N. Mukherjee, R. N. Zare, *Science* **374**, 960 (2021).
4. J. Onvlee *et al.*, *Nat. Chem*. **9**, 226 (2017).
5. B. Nichols *et al.*, *Chem. Sci*. **6**, 2202 (2015).
6. D. Yuan *et al.*, *Nat. Chem*. **10**, 653 (2018).
7. D. Dai *et al.*, *Science* **300**, 1730 (2003).
8. J. Zhang *et al.*, *Phys. Rev. Lett*. **96**, 093201 (2006).
9. D. Yuan *et al.*, *Science* **362**, 1289 (2018).
10. F. J. Aoiz, *Science* **368**, 706 (2020).

### ACKNOWLEDGMENTS

We acknowledge support from the National Natural Science Foundation of China (grant nos. 21688102 and 22125302), the Chinese Academy of Sciences (grant no. XDB17010000), the Guangdong Science and Technology Program (grant nos. 2019JC01X091 and 2019ZT08L455), and the Shenzhen Science and Technology Program (grant no. ZDSYS20200421111001787).

10.1126/science.abm5536

METALLURGY

# Strength through high slip-plane density

## Subjecting a multicomponent alloy to cyclic torsion can create a strong, ductile material

By **Jien-Wei Yeh**

Multicomponent or high-entropy alloys (HEAs), in which no one metal predominates, have expanded the alloy design space. Like conventional alloys, HEAs with greater strength often suffer from low ductility, which can limit their toughness—their ability to absorb energy without fracturing. On page 984 of this issue, Pan *et al.* (*1*) propose a methodology to improve the strength of HEAs while maintaining ductility. They engineer gradient dislocation cell structures in a single face-centered cubic (fcc) phase $Al_{0.1}CoCrFeNi$ alloy that contains nanoscale (~50 nm) randomly oriented cells in grains. The strength and ductility of this material, as well as the yield strength (the stress needed to start plastic deformation), is superior to that of the same composition with coarse or fine grains and of other metals with gradient nanograins and nanotwins.

It may seem counterintuitive that crystal defects such as dislocations, stacking faults, and twin boundaries can impart favorable mechanical properties to metals. Ductile materials can fail if an internal crack develops during plastic deformation under an applied stress. Stress acts on any unit area in a material under an applied force, and the stresses perpendicular to the area (normal stress) and parallel to it (shear stress) coexist. The latter causes the upper portion of crystal to glide along a plane between the upper and lower portions (see the figure), and dislocations allow crystals to glide and deform under a shear stress that can be several orders of magnitude less than the theoretical shear strength of a perfect crystal (*2*).

Dislocations also allow a polycrystalline metal to be ductile because the crystal glide in dislocation-free grains would actually be prone to cause fracture at grain boundaries (see the figure). Thus, dislocations allow metals to be shaped into various forms (such as flattening a sheet with rolling presses) with much less stress and expended power. Metal alloys are designed to exhibit sufficient plasticity, formability, and workability to enable their processing and sufficient toughness to avoid catastrophic failures in their intended applications.

High strength is often achieved with methods that inhibit dislocation motion (such as precipitation hardening), but this approach often decreases ductility (*3*, *4*), and many research efforts have been directed to reverse this trend. For example, the strategy suggested by the Considère criterion for necking, that higher strain-hardening rates lead to a high tensile strength and uniform elongation, is useful in the development of transformation-induced plasticity alloys and twinning-induced plasticity alloys (*3*, *4*).

Another strategy is based on the formation of more dislocations along denser and more uniform active slip planes, which should increase strength and ductility while accommodating strain. In this "high-accommodation strategy," low stacking-fault energy, lattice distortion, small grain size, dislocation cell structure, nanoprecipitates, dense stacking faults, and nanotwins could enhance accommodation of higher densities of slip planes (*4*, *5*). The approach of Pan *et al.* uses such positive factors of high-accommodation strategy to provide high strength along with large ductility.

Simple slip in a dislocation-free grain of a polycrystalline metal can cause decohesion (debonding) at grain boundaries and interfaces between different phases by forming an atomic-scale step (see the figure, top). As an example, five edge dislocation slips in one slip plane can result in decohesion, but five dislocation slips dispersed in five slip planes would not cause decohesion under the same shear strain. To extend the same concept, discrete slip bands cause a higher stress concentration against obstacles to initiate cracks than the uniform case at the same shear strain (see the figure, middle).

Thus, increasing the active slip planes and dislocation density could increase both strength and ductility as shear strain proceeds during deformation. Shear strain generally increases the dislocation density by operating dislocation sources such as grain boundaries, interphase boundaries, and Frank-Read sources (*6*). The latter could be formed by a screw dislocation if it cross-slips from one plane to another plane and then cross-slips to the plane parallel to the original plane (*7*). Smaller interplanar spacing in such double cross-slips leads to denser dislocations in the grains. Low stacking-fault energy in a distorted lattice could enhance the density of active slip planes.

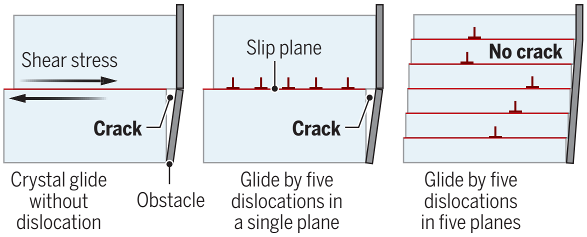
The gradient dislocation cell structure from surface to core formed by a

### Accommodating strength and toughness

Dislocations allow metals to be ductile. To make strong alloys, dislocation movement needs to be impeded, but this tends to make the materials brittle. Pan *et al*. show that mechanically introducing nanoscale dislocation cell structures creates a strong ductile alloy.

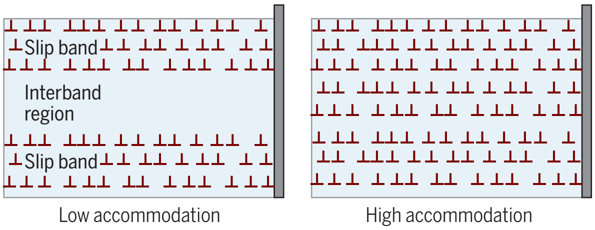
**Slipping versus cracking**

Edge dislocations (⊥) enhance the crystal lattice to slip and cause shape change, but cracking depends on their distribution.


Shear stress
Crack
Obstacle
Crystal glide without dislocation
Slip plane
Crack
Glide by five dislocations in a single plane
No crack
Glide by five dislocations in five planes

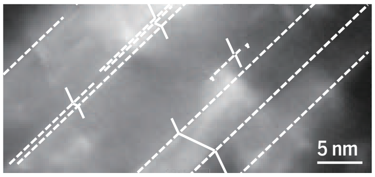
**Low versus high accommodation**

If deformation forms localized slip bands, the limited slip between bands can cause cracking at obstacles (left), unlike a dense and uniform distribution (right).



**Dense stacking faults**

An alloy processed by Pan *et al.* has ultradense stacking faults and twin boundaries (solid lines) along dense slip planes (dashed lines).



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*; (PHOTO) PAN *ET AL.* (*1*)

---

High Entropy Materials Center, Department of Materials Science and Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan. E-mail: jwyeh@mx.nthu.edu.tw

cyclic-torsion process demonstrated the variation of accommodation capacity. The highest capacity occurred in the outmost surface layer after 40% tensile strain, as verified by the dense micrometer-length stacking fault bundles through multiple dislocation cells and the highly dense stacking faults and twin boundaries in the vicinity of dislocation-cell walls (see the figure, bottom). Stacking fault probability and full-dislocation density in this layer are also highest. The progressive formation of a high density of tiny stacking faults and twins upon initial straining originates from abundant low-angle dislocation cell walls.

All of these results indicate that a high density of dislocation cell walls acting as dislocation sources, along with an effectively lower stacking fault energy for double cross-slips, produce denser active slip planes and dislocations. As a result, the evolution of this structure after yielding results in an increased strain-hardening rate, which increases both the tensile strength and uniform elongation. In addition, the gradient dislocation cell structures not only provide high yield strength but also large total elongation by activating dense slip planes that could share the plastic strain and reduce the potential for stress concentration that could induce cracks.

The work of Pan *et al.* could lead to several practical applications. The cyclic torsion process, a variation of it, or shot peening processes could be used to enhance the mechanical properties of structural parts. In addition, the implied high accommodation strategy can inspire further research on other medium-entropy (*8*) and high-entropy fcc alloys with different amounts of stacking-fault energy to achieve superior properties. Researchers should pursue further studies on the fatigue, corrosion, and stress corrosion behaviors of alloys because the free surface of a component preferentially deforms under applied stress (*9*) and is attacked more often than its interior by chemical species. Thus, more new methodologies to create high accommodation structure and superior properties for the materials can be expected. ■

### REFERENCES AND NOTES

1. Q. Pan *et al.*, *Science* **374**, 984 (2021).
2. M. A. Meyers, K. K. Chawla, in *Mechanical Metallurgy: Principles and Applications* (Prentice-Hall, 1984), pp. 193–199.
3. Y. F. Ye *et al.*, *Mater. Today* **19**, 349 (2016).
4. T. Yang *et al.*, *Science* **362**, 933 (2018).
5. M. C. Gao *et al.*, Eds., in *High-Entropy Alloys: Fundamentals and Applications* (Springer, 2016), pp. 82–97.
6. F. C. Frank, W. T. Read Jr., *Phys. Rev.* **79**, 722 (1950).
7. M. A. Meyers, K. K. Chawla, in *Mechanical Metallurgy: Principles and Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1984), pp. 257–264.
8. J. Ding, Q. Yu, M. Asta, R. O. Ritchie, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8919 (2018).
9. M. R. Stoudt, R. C. Cammarata, R. E. Ricker, *Scr. Mater*. **43**, 491 (2000).

10.1126/science.abm0120

---

**EPIDEMIOLOGY**

# Untangling the evolution of dengue viruses

## The push and pull of dengue virus serotype evolution influences epidemic potential

*By* **Pejman Rohani**[1] *and* **John M. Drake**[2]

The dengue virus (DENV) is composed of four genetically distinct clades regarded as different serotypes, DENV1 to DENV4, owing to their distinct interactions with antibodies in human blood serum (*1*). DENV is vector-borne, transmitted by *Aedes* mosquitoes, and it infects an estimated 400 million people every year, primarily in Southeast Asia and South America (*2*). Infection with DENV can result in either mild or severe disease. What makes the epidemiology and evolution of DENV fascinating is the importance of whether a repeat infection is with the same ("homotypic") or a different ("heterotypic") serotype. An open question is how interactions among DENV serotypes affect their evolutionary and epidemiological trajectories (*3*). On page 999 of this issue, Katzelnick *et al.* (*4*) studied antigenic evolution using ~350 DENVs isolated in Bangkok, Thailand, over a 20-year period. They find evidence of DENV serotypes evolving away from each other with implications for dengue epidemics.

It is thought that immunity to homotypic infection is lifelong (*5*), whereas heterotypic protection wanes over 1 to 2 years (*6, 7*). Moreover, heterotypic secondary infections may be substantially more severe than the primary infection because non-neutralizing, cross-reactive antibodies bind to the virus, forming an antibody-virus complex that binds to Fcγ receptors expressed on circulating monocytes and boosts viral replication in these cells. This phenomenon is called antibody-dependent enhancement (ADE) (*8*). Although the consequences of ADE for disease severity have been extensively studied (*3, 5, 7*), its impact on antigenic evolution remains poorly understood.

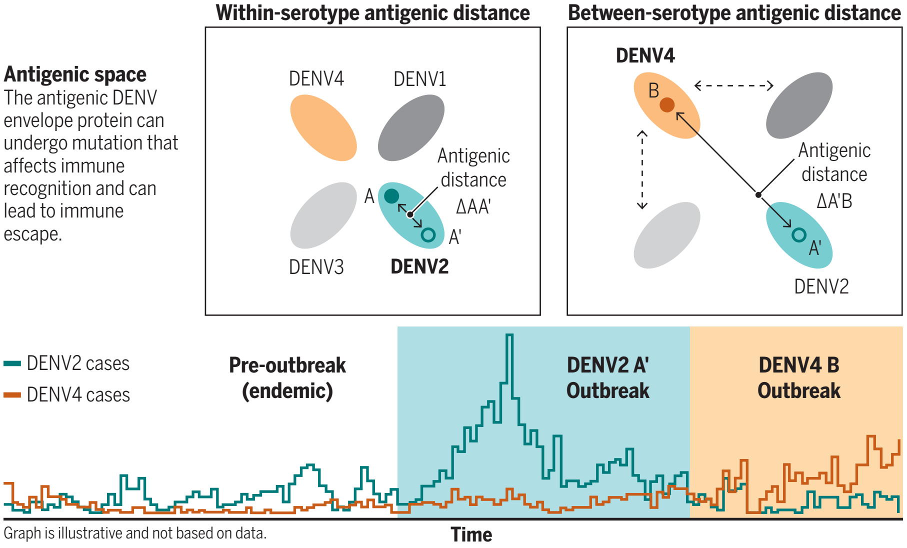
Katzelnick *et al.* found clear evidence for patterns of antigenic evolution both within and between DENV serotypes. These findings were based on painstaking antigenic characterization of the envelope protein, using a plaque reduction neutralization test (PRNT) that measures how well serum drawn from an individual infected with one virus neutralizes another virus in vitro. Data from these assays enabled the construction of a matrix of pairwise neutralization titers—antigenic "distances"—that were used to map antigenic similarity among virus isolates (*9*). The authors characterized the path through antigenic space taken by each serotype as it evolved by looking at the fluctuations in these relative distances.

Notably, viruses exhibited substantial antigenic variability when compared with other DENVs within the same serotype, with many isolates being more similar to viruses of a different serotype than to the ancestral viruses of the same serotype, consistent with prior findings (*10*). Despite this within-serotype antigenic diversity, antigenic distance from the ancestral viruses in 1994 and 1995 tended to increase through time, indicating that, on average, viruses were becoming more antigenically distinct. So, how might antigenic evolution affect virus transmission and epidemiology? In the case of seasonal influenza viruses, antigenic mismatch between circulating viruses and the prevailing immunity profile of their human hosts, either from recent infection or vaccination, gives rise to large seasonal outbreaks (*11, 12*). To date, there has been little information about how this might work in the context of DENV. To examine the epidemiological consequences of this antigenic tug-of-war, Katzelnick *et al.* compared the annual incidence of infections with each DENV serotype with the pairwise within-serotype antigenic distance in the same year. They found that antigenic novelty (large distances) correlates with outbreak size (see the figure). One possible explanation for this phenomenon is the evasion of homotypic immunity.

Katzelnick *et al.* also examined the evolution of antigenic distance among serotypes, where novelty was also found to be advantageous, with three particular sero-

---

[1] Odum School of Ecology, Department of Infectious Diseases, College of Veterinary Medicine, the Center for the Ecology of Infectious Diseases and the Center for Influenza Disease and Emergence Research (CIDER), University of Georgia, Athens, GA, USA. [2] Odum School of Ecology, the Center for the Ecology of Infectious Diseases and CIDER, University of Georgia, Athens, GA, USA.
Email: rohani@uga.edu


Within-serotype antigenic distance
Between-serotype antigenic distance
Antigenic space
The antigenic DENV envelope protein can undergo mutation that affects immune recognition and can lead to immune escape.
DENV4
DENV1
Antigenic distance
ΔAA'
A
A'
DENV3
DENV2
DENV4
B
Antigenic distance
ΔA'B
A'
DENV2
DENV2 cases
DENV4 cases
Pre-outbreak (endemic)
DENV2 A' Outbreak
DENV4 B Outbreak
Time

## The relationship between antigenic distance and outbreaks

Antigenic evolution of dengue virus (DENV) influences susceptibility to infection. Compared to a 1994/1995 DENV2 strain (A), the antigenic distance of a related DENV2 variant (A') becomes large enough to evade homotypic immunity, resulting in an outbreak. Subsequently, the antigenic distance of DENV4 from other serotypes increases enough to overcome heterotypic immunity, and an outbreak occurs.

Graph is illustrative and not based on data.

types (DENV1, DENV3, and DENV4) generally moving away from each other during this period, leading to a 65% reduction in serum neutralization. However, they also observed an interesting nonlinear pattern of antigenic change through time, with outbreaks of each serotype leading to reduced antigenic distances between serotypes. The authors speculate that this may be due to the effects of selective sweeps following an epidemic—or perhaps constraints on viral evolution imposed either by fitness in the mosquito vector or by the structure of the envelope protein. In turn, there is a complex relationship between epidemic size and pairwise between-serotype antigenic distance: In DENV1, the association between antigenic distance with other serotypes and epidemic size is strongly negative, whereas for DENV3 it is strongly positive. Why this is the case remains to be explained.

The authors investigated the evolution of DENVs relative to a reference point, chosen to be the center of the antigenic map, which may perhaps provide some indication about the structural constraints on DENV evolution. Over time, all DENVs exhibited a tendency to move away from the center of the map (resulting in an overall 40% reduction in neutralization over 20 years). Despite this general trend, however, these antigenic dynamics tended to oscillate and are only weakly negatively correlated with epidemic size. This means that there must be additional yet-unidentified factors that are driving antigenic evolution. The oscillating tendency is a clue and should be investigated more closely.

There are at least two promising areas for future research. One is to use a combination of challenge experiments, cell culture assays, and mosquito inoculations to better characterize the structural changes in the envelope protein that determine viral fitness. The other, multiscale phylodynamic modeling that simultaneously considers viral evolution and transmission dynamics, will be essential to unraveling the cause-and-effect relationships between outbreaks and antigenic trajectories (*13, 14*). It has been 6 years since the Dengvaxia vaccine was first licensed. Given the propensity of DENVs to evolve, it is important to know when it is time for an updated vaccine that uses new antigens. Accordingly, a more complete understanding of the processes involved in antigenic evolution could be of great public health benefit. ■

### REFERENCES AND NOTES

1. R. S. Lanciotti, D. J. Gubler, D. W. Trent, *J. Gen. Virol.* **78**, 2279 (1997).
2. S. Bhatt *et al.*, *Nature* **496**, 504 (2013).
3. S. M. Bell, L. Katzelnick, T. Bedford, *eLife* **8**, e42496 (2019).
4. L. C. Katzelnick *et al.*, *Science* **374**, 999 (2021).
5. A. B. Sabin, *Am. J. Trop. Med. Hyg.* **1**, 30 (1952).
6. N. G. Reich *et al.*, *J. R. Soc. Interface* **10**, 20130414 (2013).
7. H. Salje *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9535 (2012).
8. S. B. Halstead, P. Simasthien, *Yale J. Biol. Med.* **42**, 276 (1970).
9. D. J. Smith *et al.*, *Science* **305**, 371 (2004).
10. L. C. Katzelnick *et al.*, *Science* **349**, 1338 (2015).
11. K. Koelle, S. Cobey, B. Grenfell, M. Pascual, *Science* **314**, 1898 (2006).
12. D. M. Skowronski *et al.*, *Clin. Infect. Dis.* **68**, 1754 (2019).
13. B. T. Grenfell *et al.*, *Science* **303**, 327 (2004).
14. E. M. Volz, I. Siveroni, *PLOS Comput. Biol.* **14**, e1006546 (2018).

10.1126/science.abm6812

---

DEVELOPMENT

# Digging into X chromosome inactivation

## Characterizing X chromosome inactivation in nonhuman primates reveals some surprises

*By* **Edith Heard**[1] *and* **Claire Rougeulle**[2]

Dosage matters during development. Even a slight over- or underexpression of some genes can lead to developmental abnormalities. Sex chromosomes pose a particular challenge when it comes to gene dosage differences between the sexes. Equalizing X-linked gene dosage between sexes through X chromosome inactivation (XCI), in which one X chromosome in females is transcriptionally silenced, is a hallmark of mammalian species and is considered essential for female development. On page 954 of this issue, Okamoto *et al.* (*1*) shake the assumption that XCI is always a precocious event and unveil, with great precision, the rather late and varied establishment of XCI during cynomolgus macaque development. They also reveal the timing of XCI reversion in the female germ line, in one of the most complete studies in a primate. Their study highlights the remarkable diversity in XCI features between mammalian species and forces us to revisit some fundamental questions about why and how dosage compensation is achieved.

The dynamics, mechanisms, and molecular actors of XCI have been most closely scrutinized in mice, in which XCI happens early on in preimplantation development [starting at the 4- to 8-cell stage, around 2 days of embryogenesis (E2)], triggered by expression of the X-inactive specific transcript (XIST) long noncoding RNA. XIST RNA coats the X chromosome in cis and triggers gene silencing as well as chromatin modifications through mechanisms that are only now being unraveled. Achieving appropriate sex-linked gene dosage was assumed

---

[1]Director's Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [2]Department of Epigenetics and Cell Fate, Université de Paris, CNRS, F-75006 Paris, France. Email: edith.heard@embl.org; claire.rougeulle@u-paris.fr

GRAPHIC: K. FRANKLIN/*SCIENCE*

to be an early developmental requirement, and the absence of XCI (for example, when *Xist* is deleted) leads to early postimplantation lethality of female mouse embryos.

By combining single-cell X-linked gene expression analyses with chromatin profiling, Okamoto *et al.* find that early cynomolgus embryo development proceeds without any overt signs of XCI. Notably, it is only after implantation, at E9, that one of the two female X chromosomes is silenced, with different lineage-specific kinetics. XCI occurs concomitantly to, or soon after, implantation in trophectoderm-derived extraembryonic lineages, whereas it is not completed before E17 in lineages derived from the inner cell mass, which contribute to the embryo proper. This order of events differs substantially from that in the mouse. Previous studies that used cultured preimplantation human and rabbit embryos had provided a first glimpse into the diversity of XCI strategies across mammals (*2*, *3*), but capturing XCI in its entirety during development of a nonrodent species has not been accomplished before.

Cynomolgus XCI has more similarities to what little is known about human XCI but also shows some differences (see the figure). Similar to human embryos (but unlike mice), accumulation of XIST RNA occurs on both active X chromosomes for several days in female cynomolgus embryos (and on the single X chromosome in males). XIST accumulation and XCI are transiently uncoupled, for a period of up to 2 weeks in some cells. Nonhuman primate, and probably human, XCI therefore involve a switch not only in XIST expression but also in XIST's ability to induce XCI. In contrast to humans, however, cynomolgus XIST accumulation is accompanied by X chromosome enrichment with the Polycomb repressive complex (PRC)–associated chromatin marks histone H2A $\text{Lys}^{119}$ ubiquitylation (H2AK119Ub) and histone H3 $\text{Lys}^{27}$ trimethylation (H3K27me3). This reinforces the latest models for XCI in which PRC-mediated marks are not sufficient for global establishment of X-linked gene silencing but rather participate in spreading and locking-in of the inactive state (*4*).

Many matters remain unanswered: what regions of the X chromosome are initially targeted by XIST RNA; why XIST RNA does not immediately induce gene silencing, through the repressor protein SPEN, as in mice (*5*); how the X chromosomes remain active despite the overall addition of repressive PRC-mediated chromatin marks; and whether early enrichment for H2AK119Ub and H3K27me3 is required for future XCI. The study by Okamoto *et al.* touches on the nature of the mechanisms that prevent the recruitment of PRCs by XIST at the equivalent stage in humans, although some studies that used cellular models for human preimplantation development have also reported the enrichment of H3K27me3 on XIST-coated active X chromosomes (*6*).

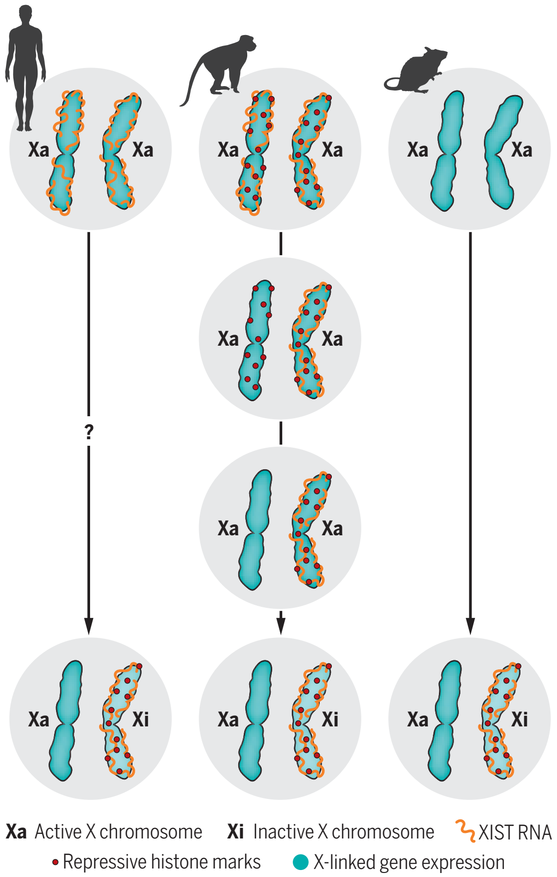
The study by Okamoto *et al.* captures the transition from before to after XCI in nonhuman primates in vivo. Although XIST initially accumulates on both the future active and inactive X chromosomes, *XIST* is repressed on the future active X chromosome (Xa) before gene silencing is triggered on the other X. Intriguingly, H3K27me3 and H2AK119Ub enrichment persist longer than XIST RNA coating on the future Xa. The dynamics observed are consistent with the hypothesis of a switch in XIST's ability to induce silencing during early embryonic development, preventing inactivation of both X chromosomes. This transition does not appear to involve an intermediate state with no XIST expression, as seen in humans by use of cellular models (*6*).

The rather late establishment of XCI raises the question of what happens in the germ line. Primordial germ cells (PGCs), the founding population of the germ-cell lineage, undergo global epigenetic resetting, which is thought to include X chromosome reactivation (XCR). In cynomolgus macaques, PGCs emerge from E11 to E17, concomitantly with or before XCI establishment. From the study of Okamoto *et al.*, it appears that most cells enter the germline specification process in a pre-XCI state and therefore altogether escape the XCI cycle, a situation that has never previously been reported.

That primate development proceeds for up to 17 days without systematic XCI is particularly intriguing and urges reconsideration of the links between XCI and differentiation and the time and extent to which gene dosage compensation is needed. Dosage compensation may take different forms. In humans, the occurrence of a transient mode of dosage compensation, with expression from both X chromosomes being "dampened" in females before XCI, is currently debated (*6*–*9*). The Okamoto *et al.* study could be consistent with X chromosome dampening occurring transiently in females only, independently of XIST and PRC-mediated chromatin marks. Another form of dosage compensation involves up-regulation of gene expression from the Xa chromosomes in both sexes, to equalize sex chromosome expression with that of autosomes. The in vivo analysis by Okamoto *et al.* is coherent with Xa activity up-regulation occurring from E6 in males. In females, X chromosome dampening may begin before XCI and Xa activity up-regulation, but the exact order and nature of events requires further clarification. ■

## Switching off the X chromosome

Okamoto *et al.* report the kinetics of events leading to X chromosome inactivation in cynomolgus macaque embryogenesis. The activity status, the accumulation of X-inactive specific transcript (XIST) RNA, and the presence of repressive histone marks, in comparison with mice and human X chromosome inactivation, are indicated.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

### REFERENCES AND NOTES

1. I. Okamoto *et al.*, *Science* **374**, eabd8887 (2021).
2. I. Okamoto *et al.*, *Nature* **472**, 370 (2011).
3. C. Vallot *et al.*, *Cell Stem Cell* **20**, 102 (2017).
4. A. Bousard *et al.*, *EMBO Rep.* **20**, e48019 (2019).
5. F. Dossin *et al.*, *Nature* **578**, 455 (2020).
6. A. Sahakyan *et al.*, *Cell Stem Cell* **20**, 87 (2017).
7. S. Mandal *et al.*, *Stem Cell Reports* **14**, 745 (2020).
8. J. C. Moreira de Mello, G. R. Fernandes, M. D. Vibranovski, L. V. Pereira, *Sci. Rep.* **7**, 10794 (2017).
9. S. Petropoulos *et al.*, *Cell* **165**, 1012 (2016).

10.1126/science.abm1857

Skiers Marcello Cominetti and Giulia Monego traverse a section of the Italian Alps.

CLIMATE CHANGE

# Shrinking snowpacks, anxious snow lovers

## A cold-weather enthusiast confronts connections between the cryosphere and human culture

*By* **Mark Serreze**

Our planet's shrinking cryosphere—its frozen water in all its different forms—is a particularly troubling aspect of climate change. Thinning mountain snowpacks, rising flood risks, ecosystem disturbances and habitat loss, rising sea levels, the growing geopolitical importance of the Arctic, and threats to winter recreation are all tied to the ongoing loss of our planet's ice.

As a climate scientist who has studied the north polar region for nearly 40 years, I am all too familiar with what is happening to our planet's ice, but much of my research has focused on the physics and quantitative observables of climate change. After reading *The Last Winter* by Porter Fox, I have come to realize that my knowledge of the cryosphere and its importance to our planet, while perhaps deep, has also been rather narrow.

*The Last Winter* chronicles the myriad impacts of the loss of snow and ice through the tales of ranchers, scientists, adventurers, hunters, and guides, as well as people who simply love winter. The book begins in a seemingly odd place—the middle of a raging wildfire in the North Cascade mountains, where a local rancher is watching all that he loves go up in smoke—but it succeeds in setting the stakes posed by our ever-shrinking frozen water reserves.

From here, Fox's narrative expands his scope, with stories about the shrinking snowpack and glaciers of the western United States, anecdotes from scientists and students working on the Juneau Icefield of Alaska, and tales from the rapidly warming European Alps and the melting Greenland Ice Sheet. Together, such stories illustrate how connections between the cryosphere, Earth's various ecosystems, and human economies and culture are complex, steeped in history, and filled with unknowns.

THE LAST WINTER
PORTER FOX

**The Last Winter**
*Porter Fox*
Little, Brown, 2021.
320 pp.

Fox grew up in Maine (as did I), where even today, one can still find "real winters" and revel in them. Throughout the book, he weaves in tales of how snow and ice have shaped human culture and history, from serving as primal forces that humanity sought to conquer—as when European adventurers attempted to connect the Atlantic and Pacific oceans via a navigable sea route through the Arctic Ocean—to being conceived of as resources to be exploited, to being sources of joy and recreation. If the Alps lose their glaciers and their snow, he wonders, with regard to the latter issue, do they also lose a crucial aspect of their identity—and with it a culture that has endured for centuries?

Fox recognizes that to understand the climate system, one must appreciate not just the intimate connections between the oceans, atmosphere, and land but also the importance of snow and ice in making these connections. Changes in snow cover in Siberia, for example, can, by affecting the albedo (reflectivity) of the surface, influence weather patterns in the United States. Fox also offers readers many relevant facts and figures, such as how quickly mountain snowpacks are shrinking and when the glaciers of the Cascades might disappear, as well as explanations of the physics underlying these changes.

Although Fox is a journalist, not a climate scientist, he has done his homework on this subject, including deep dives into paleoclimate and archaeological studies that put our present predicament into perspective. Ice played key roles in the evolution and migratory patterns of our species, he reveals. During the last ice age, for example, lower sea levels enabled the peopling of North America, and the forests of East Africa dried up, forcing fundamental changes in hunting and gathering practices (*1*, *2*).

*The Last Winter* is occasionally bogged down by Fox's digressions into skiing and his sometimes fawning descriptions of the scientists we meet along the way. (I would not have described the late Koni Steffen as "the father of Greenland ice data," for example.) The book is also arguably incomplete. Some stories and perspectives from the rapidly warming Himalayas, for example, where glaciers are shrinking and water resources are threatened, might have added a new dimension to the narrative. But these are minor qualms.

Ultimately, *The Last Winter* is a welcome addition to the growing genre of books seeking to link human faces with global climate change. When telling such tales, it can be difficult to find the right balance between science and storytelling. *The Last Winter* hits the mark. ■

---

The reviewer is at the National Snow and Ice Data Center (NSIDC), University of Colorado, Boulder, CO 80309, USA, and the author of *Brave New Arctic: The Untold Story of the Melting North* (Princeton Univ. Press, 2018). Email: mark.serreze@colorado.edu

REFERENCES AND NOTES

1. M. A. Kelly *et al.*, *Geology* **42**, 519 (2014).
2. T. Goebel, M. R. Waters, D. H. O'Rourke, *Science* **319**, 1497 (2008).

10.1126/science.abm3456

PHOTO: DAVID REDDICK

ANTHROPOLOGY

# Incomplete Indigenous landscapes

## A new text confronts how bias tainted efforts to preserve native cultures of the Americas

*By* **Paulette Steeves**

Non-Indigenous people's ideas about Indigenous communities of the Americas have been broadly shaped by Euro-American anthropological perspectives that frequently misrepresent Indigenous culture and history. In his new book, *Prophets and Ghosts*, Samuel Redman traces the history of Euro-American "salvage anthropology"—the widespread effort to document the culture and customs of Indigenous societies of the Americas before their (presumed) imminent extinction—from its early beginnings in the 19th century to recent times.

Anthropologist Frank Hamilton Cushing accompanies a Zuni party to meet US President Chester Arthur in 1882.

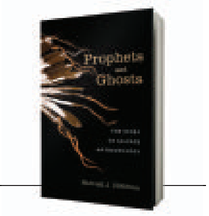
The idea of stockpiling Indigenous cultural information was built on a common assumption in the mid-1800s that the Indians of the Americas were fated to become extinct. As early as the 18th century, notes Redman, "people in the Americas…pondered whether or not American Indians were a doomed race and what could be done to preserve some aspect of their vanishing societies." During this period, Indigenous societies were considered primitive by many anthropologists of European descent. This racist view aligned with commonly held notions within Euro-American society.

Redman traces the early development of salvage anthropology in the Americas; introduces major players, including Lewis Henry Morgan, Henry Schoolcraft, and Ely Parker; and recounts how collectors amassed an empire of antiquities that belonged to supposedly vanishing races. So vast were these collections of artifacts and documentation of languages and customs that between 1860 and 1940 the number of museums in the United States almost doubled, he notes.

Salvage anthropology, which relied on securing philanthropic contributions to fund fieldwork and establishing relationships with Indigenous communities, frequently led to biases with regard to which artifacts and traditions were collected and from which communities. The ensuing publications and museum exhibits provided an incomplete picture of America's Indigenous landscapes.

Readers of Redman's account learn how some salvage anthropologists grew into an understanding of Indigenous people as gifted intellectuals. Ethnographer and "Indian agent" (an individual authorized to interact with native people on behalf of the US federal government) Henry Schoolcraft, for example, initially held Indigenous people in low regard, "describing them as slothful, amoral savages," yet after taking the time to learn the language of the Ojibwa and getting to know the people he studied, his opinion slowly changed, and he eventually married the daughter of a prominent Ojibwa leader.

**Prophets and Ghosts: The Story of Salvage Anthropology**
*Samuel J. Redman*
Harvard University Press, 2021. 328 pp.



Not all anthropologists of this time were open to challenging their biased worldviews, however. For instance, Aleš Hrdlička, a Smithsonian archaeologist, maintained a hard stance on racial purity, believing that intertribal marriage would lead to a loss of "full-bloods" to study.

The efforts of Euro-American salvage anthropologists of this era were not entirely unwarranted, misguided as their execution often was. The lifeways, cultural practices, and homelands of many Indigenous communities were indeed being decimated at this time. Redman's narrative paints a vivid portrait of the dedicated work done by linguists, artists, painters, and filmmakers who sought to preserve the history of Indigenous cultures. The linguist John Peabody Harrington, for example, recorded nearly a million pages of linguistic and ethnographic data on various Indigenous communities, notes Redman.

The book also discusses contemporary Indigenous activists and their work to reclaim and repatriate histories, stories, ancestors, artifacts, and lands, ending with a brief discussion of the Native American Graves Protection and Repatriation Act (NAGPRA). NAGPRA came into law in 1990 and required institutions holding Native American remains and artifacts of cultural and religious patrimony to create a listing of their collections, to make these lists available to Native American communities, and to repatriate items to communities that claimed them. "The legal framework for returning these remains and sacred materials to Indigenous communities further extends their recognition as sovereign nations," notes Redman.

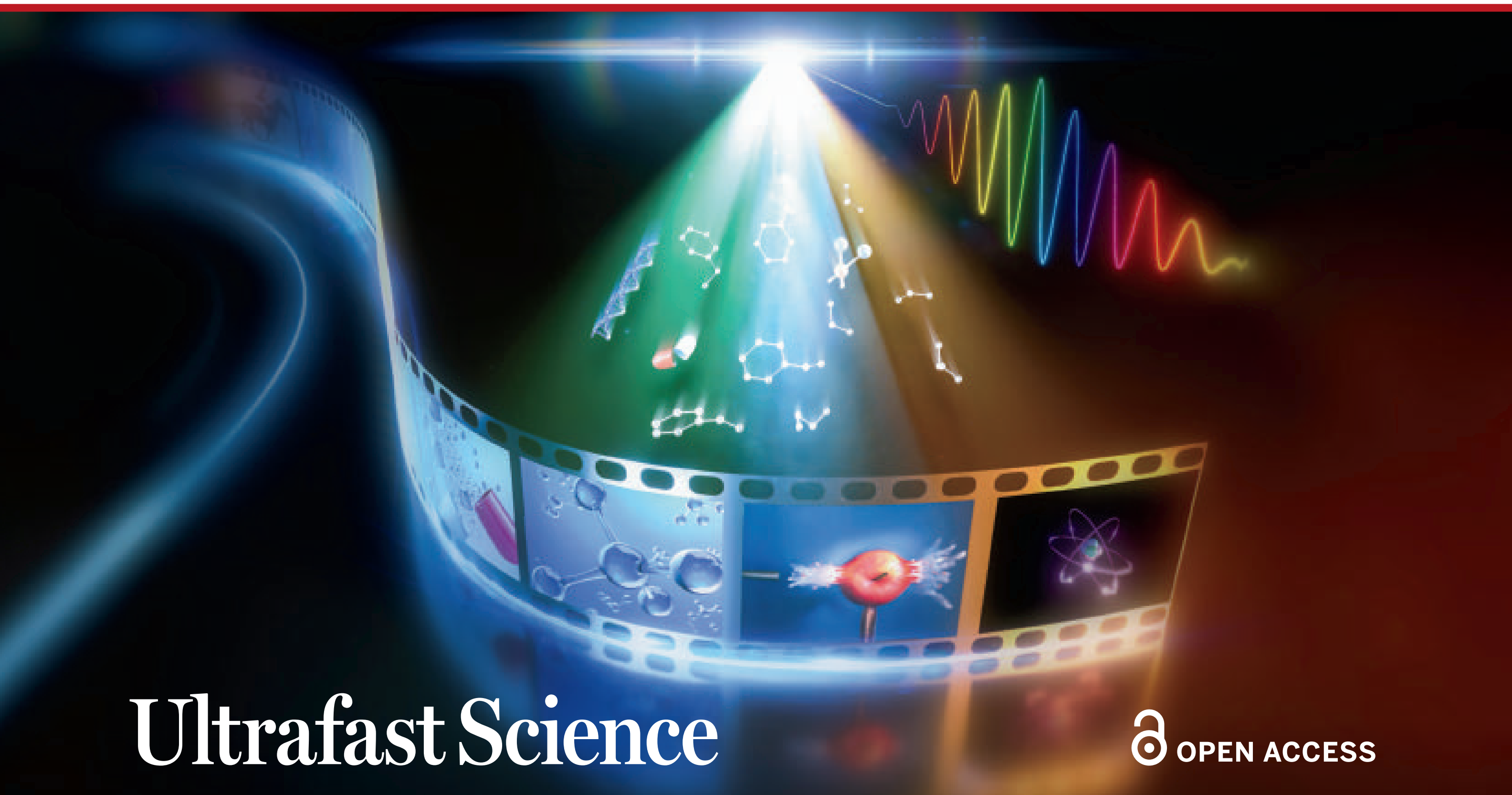
Scholars today are increasingly heeding the challenge to "decolonize" their worldviews in an effort to understand how colonization and racism have informed long-held and often erroneous beliefs about Indigenous populations. *Prophets and Ghosts* is a welcome addition to this effort and is a must-read for anyone seeking to confront racist worldviews and make the world a better place for all. ■

10.1126/science.abm5544

---

The reviewer is the Canada Research Chair Tier II in Indigenous History Healing and Reconciliation at Algoma University, Sault Ste. Marie, ON P6A 2G4, Canada. Email: paulette.steeves@algomau.ca

PHOTO: NATIONAL PORTRAIT GALLERY/SMITHSONIAN INSTITUTION



CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS

Ultrafast Science

OPEN ACCESS

*Ultrafast Science* is an online-only, Open Access journal published in affiliation with **Xi'an Institute of Optics and Precision Mechanics (XIOPM)** of CAS and distributed by the **American Association for the Advancement of Science (AAAS)**. The mission of the journal is to build a platform for ultrafast scientific research; present new theories, concepts, ideas, technologies and progress in this field; publish the latest research representing the forefront of the discipline and that of international concern; and promote academic exchange and development worldwide. *Ultrafast Science* will publish high-quality original research articles, comprehensive reviews, editorials, and perspectives which feature high novelty, significance and technical quality in cutting-edge and emerging topics in ultrafast science with broad interest from scientific communities.

**Submit your research to *Ultrafast Science* today!**
Learn more at **spj.sciencemag.org/ultrafastscience**

The Science Partner Journal (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only, Open-Access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS furthers its mission to communicate science broadly and for the benefit of all people by providing top-tier international research organizations with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.

Visit us at **spj.sciencemag.org**

**@SPJournals** **@SPJournals**

ARTICLE PROCESSING CHARGES WAIVED UNTIL OCTOBER 2024

LETTERS

Cougars (*Puma concolor*) have recently been seen in populated areas of Argentina, but attacks on people are rare.

PHOTO: FRANCO BUCCI/FUNDACIÓN REWILDING ARGENTINA

*Edited by* **Jennifer Sills**

# Minimizing fear of wildlife in urban areas

The likelihood of an increase in conflict between humans and wildlife as climate change intensifies is undeniable ("Human-wildlife conflict under climate change," B. Abrahms, Perspective, 30 July, p. 484). Therefore, the role that human fear plays in these events must be considered. The reaction of humans to animals, especially in cities, is often disproportionate to the danger they pose.

Recent videos show cougars (*Puma concolor*) in towns surrounded by natural areas in Argentina [e.g., (*1*)]. Although no cougar-human encounter occurred, newspaper headlines declared "panic" (*2*). During such events, people often assume that the potential danger to human security warrants urgent preventative—including lethal—actions. Human safety is a priority, but measures must be commensurate with the danger posed by the animal.

The fear generated by wildlife sightings does not necessarily scale with the actual threat. The probability of cougars attacking humans is extremely low. In the United States and Canada, fewer than a dozen fatal cougar attacks on people occurred over 100 years (*3*). Similarly, in Argentina, where the species is widespread and relatively abundant, only occasional attacks on people have been documented (*4*). In contrast, 109 humans died due to domestic dog attacks between 1989 and 1994 in the United States alone (*5*). Yet people live surrounded by dogs, almost fearlessly. To put the number of deaths caused by wildlife in perspective, many more people have died taking a selfie in the past decade (*6*) than as a result of an attack by a cougar.

This fear of wildlife, despite the facts, is likely a result of cognitive and social bias (*7*), a lack of knowledge about wild species and their behavior, and the negative framing used by media to present wildlife news (*8*). If we are to pursue the vision proposed by international agreements of living in harmony with nature by 2050 (*9*), we must educate humans about how to act in the case of a wildlife encounter. To minimize conflict, humans should avoid feeding wildlife, manage waste and landfills efficiently to reduce access to food resources, and keep pets responsibly to prevent predation. In the event of an encounter, humans should be prepared to use nonlethal deterrence methods. School curricula should foster the value of living in harmony with nature from a young age. To facilitate shared, safe spaces, humans can plan wildlife-friendly urban areas (*10*). The media should also be held responsible for their presentation of wildlife; news coverage can contribute to rather than interfere with the development of healthy human-wildlife relationships. Given that humans and wildlife are likely to come into contact with increasing frequency, we should take steps to prevent persecution of native wildlife species and prepare humans to accept that in most cases it is safe to share their home without fear.

**Sergio A. Lambertucci*, Pablo Plaza, Karina Speziale**
Grupo de Investigaciones en Biología de la Conservación, Instituto de Investigaciones en Biodiversidad y Medio ambiente (Consejo Nacional de Investigaciones Científicas y Técnicas–Universidad Nacional del Comahue), Quintral 1250, Bariloche, Argentina.
*Corresponding author.
Email: slambertucci@comahue-conicet.gob.ar

## REFERENCES AND NOTES

1. Diario Rio Negro, Buscan a un puma en el centro de Bariloche (2021); www.youtube.com/watch?v=SzKxTtPFx7c.
2. "Pánico en General Rodríguez por un puma suelto," Infobae (2021); www.infobae.com/sociedad/2021/09/23/video-panico-en-general-rodriguez-por-un-puma-suelto/ [in Spanish].
3. P. Beier, *Wildl. Societ. Bull*. **19**, 403 (1991).
4. C. M. Portelli, C. R. Eveling, J. Lamas, P. J. Mamaní, *Cuadernos de Medicina Forense* **18**, 139 (2012) [in Spanish].
5. J. J. Sacks, R. Lockwood, J. Hornreicht, R. W. Sattin, *Pediatrics* **97**, 891 (1996).
6. B. Weiler, A. M. Gstaettner, P. Scherrer, *Tour. Manag. Perspect*. **37**, 100778 (2021).
7. J. Cinner, *Science* **362**, 889 (2018).
8. F. Ballejo, P. Plaza, S. A. Lambertucci, *Sci. Rep*. **11**, 13512 (2021).
9. "Long-term strategic directions to the 2050 vision for biodiversity, approaches to living in harmony with nature and preparation for the post-2020 global biodiversity framework" (Convention on Biological Diversity, 2018); www.cbd.int/doc/c/0b54/1750/607267ea9109b52b750314a0/cop-14-09-en.pdf.
10. B. Apfelbeck *et al*., *Landsc. Urban Plan*. **200**, 103817 (2020).

10.1126/science.abm6560

# Surviving as a young scientist in Brazil

In recent years, Brazilian science has been hit by repeated budget cuts (*1*–*5*), including one in October that reduced a proposal for supplemental science funding by 87% (*4*). These cuts threaten ongoing and future science and technology projects, including scholarship funding for early-career researchers. Despite precarious funding for science compared with high-income countries, in 2020, Brazil was ranked 11th among nations that produced knowledge to fight the COVID-19 pandemic (*6*). Brazil's meaningful contributions are driven by early career researchers (*5*), but without proper investment, these young scientists will have no choice but to abandon their academic career or leave the country (*5*).

In Brazil, most postgraduate programs require full dedication as a condition of employment—the scholarship cannot be paired with any other source of income. Early career researchers in master's and PhD programs earn, respectively, a scholarship of US$300 and US$440 monthly (*7*), and these amounts have not been adjusted since 2013 (*7*, *8*). In that time, rising prices for products and services decreased the purchasing power of scholarships by more than 60% (*9*). Most students depend exclusively on this monthly payment to maintain their food, health, housing, clothing, and transportation. In many cases, they are also supporting their families.

Investment in training scientists is paramount to generate cutting-edge science for society's development and well-being (*10*). Thus, as Brazilian early career researchers, we demand adequate financial support from decision-makers in Brazil. If Brazil does not immediately reevaluate its budget for science and technology, the country risks losing or alienating an entire generation of Brazilian scientists.

**Celso H. L. Silva Junior[1,2,3,4]*, Yhasmin M. Moura[5,6,4], Ana C. M. Pessôa[1,2], Diego P. Trevisan[7,8,4], Flávia S. Mendes[4], João B. C. Reis[1,9], Michelle C. A. Picoli[10,4], Natalia C. Wiederkehr[2,4], Nathália S. Carvalho[1,2], Ricardo Dalagnol[1,2,7,4], Tahisa N. Kuck[11,4], Thais M. Rosan[12], Thiago S. F. Silva[13,4], Veraldo Liesenberg[14], Polyanna C. Bispo[7,4]**

[1]Tropical Ecosystems and Environmental Sciences Laboratory, São José dos Campos, SP, Brazil. [2]Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brazil. [3]Universidade Estadual do Maranhão, São Luís, MA, Brazil. [4]Remote Sensing Applied to Tropical Environments Group, Manchester, UK. [5]Institute of Geography and Geoecology, Karlsruhe Institute of Technology, Karlsruhe, Germany. [6]Centre for Landscape and Climate Research, School of Geography, Geology, and the Environment, University of Leicester, Leicester, UK. [7]Department of Geography, School of Environment, Education and Development, University of Manchester, Manchester, UK. [8]Department of Environmental Sciences, Federal University of São Carlos, São Carlos, SP, Brazil. [9]Centro Nacional de Monitoramento e Alertas de Desastres Naturais, São José dos Campos, SP, Brazil. [10]Earth and Life Institute, Georges Lemaître Centre for Earth and Climate Research, Université Catholique de Louvain, Louvain-La-Neuve, Belgium. [11]Institute for Advanced Studies, São José dos Campos, SP, Brazil. [12]College of Life and Environmental Sciences, University of Exeter, Exeter, UK. [13]Biological and Environmental Sciences, University of Stirling, Stirling, UK. [14]Department of Forest Engineering, Santa Catarina State University, Lages, SC, Brazil.

*Corresponding author. Email: celsohlsj@gmail.com

The views in the Letter are the authors' and do not necessarily represent the views of their institutions.

## REFERENCES AND NOTES

1. H. Escobar, *Science* **364**, 111 (2019).
2. J. Tollefson, *Nature* **572**, 161 (2019).
3. Sociedade Brasileira para o Progresso da Ciência, "Entidades se manifestam contra manobra sorrateira do Ministério da Economia que retira recursos do CNPq" (2021); http://portal.sbpcnet.org.br/noticias/entidades-se-manifestam-contra-manobra-sorrateira-do-ministerio-da-economia-que-retira-recursos-do-cnpq [in Portuguese].
4. R. O. Andrade, "Congresso promove novos cortes nos recursos para a ciência," *Revista Pesquisa Fapesp* (2021); https://revistapesquisa.fapesp.br/congresso-promove-novos-cortes-nos-recursos-para-a-ciencia [in Portuguese].
5. H. Escobar, "A Diáspora," *Revista Piauí* (2021); https://piaui.folha.uol.com.br/materia/a-diaspora [in Portuguese].
6. J. Bernardes, M. Dorado, "USP está entre as 20 instituições que mais publicam sobre covid no mundo," *Jornal da USP* (2020); https://jornal.usp.br/ciencias/usp-esta-entre-as-20-instituicoes-que-mais-publicam-sobre-covid-no-mundo [in Portuguese].
7. Conselho Nacional de Desenvolvimento Científico e Tecnológico, "Tabela de Valores de Bolsas no País" (2021); www.gov.br/cnpq/pt-br/acesso-a-informacao/bolsas-e-auxilios/copy_of_modalidades/tabela-de-valores-no-pais [in Portuguese].
8. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, "Valores de bolsas" (2021); www.gov.br/capes/pt-br/acesso-a-informacao/acoes-e-programas/bolsas/prestacao-de-contas/valores-de-bolsas [in Portuguese].
9. Banco Central do Brasil, "Correção de valores" (2021); https://www3.bcb.gov.br/CALCIDADAO/publico/exibirFormCorrecaoValores.do?method=exibirFormCorrecaoValores [in Portuguese].
10. L. Massarani, Y. Castelfranchi, I. Mendes, V. O. Fagundes, I. Moreira, *Anais da Academia Brasileira de Ciências* **93**, e20200204 (2021).

10.1126/science.abm8160

---

# Scientific denialism threatens Brazil

Brazil has been facing a wave of institutionalized scientific denialism that negatively affects public policies in areas such as education, environmental preservation (*1*, *2*), and public health (*3*). Among the institutionalized setbacks are the systemic cuts to the scientific budget (*4*–*6*), including a 92% cut to the National Fund for Science and Technology Development budget in October, equivalent to US$106.3 million (*4*). Less scientific funding threatens the careers of young scientists (*7*) and the productivity of senior researchers, who will likely see a decrease in students' motivation to develop their projects. Reducing Brazil's scientific output will exacerbate the wide-ranging consequences of the government's anti-science policies.

The ideology of scientific denialism, conveyed in actions and words by the Brazilian president, Jair Bolsonaro, has led to setbacks in the country's scientific and technological advances. The politicization of science has even fueled attacks on scientists who advocated against such setbacks (*3*, *8*). The institutional rejection of science-based strategies represents a threat to conservation, public health policies, epidemiological control, sustainable economic development, traditional communities, and the global climate (*2*, *3*).

Brazil represents a cautionary tale for other countries at risk of embracing leaders who dismiss science. Brazil's Pantanal wetlands are threatened by fires (*9*), and deforestation rates in the Amazon are growing (*2*), yet the government tries to shift blame rather than use science-based strategies to protect the environment (*2*, *10*). Mismanagement of the COVID-19 pandemic has led to more than 600,000 deaths in the country (*3*, *8*), but leaders continue to encourage antivaccine movements and promote ineffective drugs (*3*). As Brazilians face water, food, and energy crises, Brazil's leadership continues to deny climate change (*2*). Although Brazil likely could not have prevented these challenges, science denialism has hobbled the government's ability to effectively address them.

Brazil plays a major role in the global economy and global climate system, in addition to housing abundant biodiversity. However, given the scientific denialism embedded in its institutions, Brazil may not be able to protect its vast resources. It is crucial to combat the science denialism contaminating the country's decision-making. The government can begin by reversing the cuts to the science budget. Investing in science and valuing scientific output will promote economic development and demonstrate a commitment to addressing the climate, public health, and biodiversity crises, which are already a global concern.

**Luisa Maria Diele-Viegas[1]*, Juliana Hipólito[2,3,4], Lucas Ferrante[4]**

[1]Departamento de Biologia, Federal University of Alagoas, Maceió, Alagoas, Brazil. [2]National Institute of Amazonian Research, Manaus, Amazonas, Brazil. [3]Instituto de Biologia, Federal University of Bahia, Salvador, Bahia, Brazil. [4]Programa de Biologia (Ecologia), National Institute of Amazonian Research, Manaus, Amazonas, Brazil.

*Corresponding author.
Email: luisa.mviegas@gmail.com

## REFERENCES AND NOTES

1. C. A. Nobre, *Nature* **574**, 455 (2019).
2. L. Ferrante, P. M. Fearnside, *Environ. Conserv.* **46**, 261 (2019).
3. L. Ferrante *et al.*, *J. Public Health Pol.* **42**, 439 (2021).
4. M. Rodrigues, *Nature* 10.1038/d41586-021-02886-9 (2021).
5. E. Mega, *Science* 10.1126/science.aaz1848 (2019).
6. R. D. O. Andrade, *Nature* **572**, 575 (2019).

7. R. Heyard *et al.*, *Scientometrics* **126**, 9333 (2021).
8. H. Escobar, *Science* **372**, 225 (2021).
9. E. R. Mega, *Nature* **586**, 20 (2020).
10. S. Hecht, *Curr. Hist.* **119**, 6 (2020).

10.1126/science.abm9933

## TECHNICAL COMMENT ABSTRACTS

### Comment on "A global environmental crisis 42,000 years ago"

**John Hawks**

Cooper *et al*. (Research Article, 19 February 2021, p. 811) propose that a weakening geomagnetic field prior to the Laschamps Excursion explains megafaunal extinctions and human cultural changes that they claim happened 42,000 years ago. However, these authors misrepresent both the data and interpretations of cited work on extinctions and human cultural changes, so the specific claims they make about extinctions and cultural changes are false.

**Full text:** dx.doi.org/10.1126/science.abh1878

### Response to Comment on "A global environmental crisis 42,000 years ago"

**Alan Cooper, Chris S. M. Turney, Jonathan Palmer, Alan Hogg, Matt McGlone, Janet Wilmshurst, Andrew M. Lorrey, Timothy J. Heaton, James M. Russell, Ken McCracken, Julien G. Anet, Eugene Rozanov, Marina Friedel, Ivo Suter, Thomas Peter, Raimund Muscheler, Florian Adolphi, Anthony Dosseto, J. Tyler Faith, Pavla Fenwick, Christopher J. Fogwill, Konrad Hughen, Matthew Lipson, Jiabo Liu, Norbert Nowaczyk, Eleanor Rainsley, Christopher Bronk Ramsey, Paolo Sebastianelli, Yassine Souilmi, Janelle Stevenson, Zoë Thomas, Raymond Tobler, Roland Zech**

Our paper about the impacts of the Laschamps Geomagnetic Excursion 42,000 years ago has provoked considerable scientific and public interest, particularly in the so-called Adams Event associated with the initial transition of the magnetic poles. Although we welcome the opportunity to discuss our new ideas, Hawks's assertions of misrepresentation are especially disappointing given his limited examination of the material.

Full text: dx.doi.org/10.1126/science.abh3655

### Comment on "A global environmental crisis 42,000 years ago"

**Andrea Picin, Stefano Benazzi, Ruth Blasco, Mateja Hajdinjak, Kristofer M. Helgen, Jean-Jacques Hublin, Jordi Rosell, Pontus Skoglund, Chris Stringer, Sahra Talamo**

Cooper *et al*. (Research Article, 19 February 2021, p. 811) propose that the Laschamps geomagnetic inversion ~42,000 years ago drove global climatic shifts, causing major behavioral changes within prehistoric groups, as well as events of human and megafaunal extinction. Other scientific studies indicate that this proposition is unproven from the current archaeological, paleoanthropological, and genetic records.

**Full text:** dx.doi.org/10.1126/science.abi8330

### Response to Comment on "A global environmental crisis 42,000 years ago"

**Alan Cooper, Chris S. M. Turney, Jonathan Palmer, Alan Hogg, Matt McGlone, Janet Wilmshurst, Andrew M. Lorrey, Timothy J. Heaton, James M. Russell, Ken McCracken, Julien G. Anet, Eugene Rozanov, Marina Friedel, Ivo Suter, Thomas Peter, Raimund Muscheler, Florian Adolphi, Anthony Dosseto, J. Tyler Faith, Pavla Fenwick, Christopher J. Fogwill, Konrad Hughen, Matthew Lipson, Jiabo Liu, Norbert Nowaczyk, Eleanor Rainsley, Christopher Bronk Ramsey, Paolo Sebastianelli, Yassine Souilmi, Janelle Stevenson, Zoë Thomas, Raymond Tobler, Roland Zech**

Our study on the exact timing and the potential climatic, environmental, and evolutionary consequences of the Laschamps Geomagnetic Excursion has generated the hypothesis that geomagnetism represents an unrecognized driver in environmental and evolutionary change. It is important for this hypothesis to be tested with new data, and encouragingly, none of the studies presented by Picin *et al.* undermine our model.

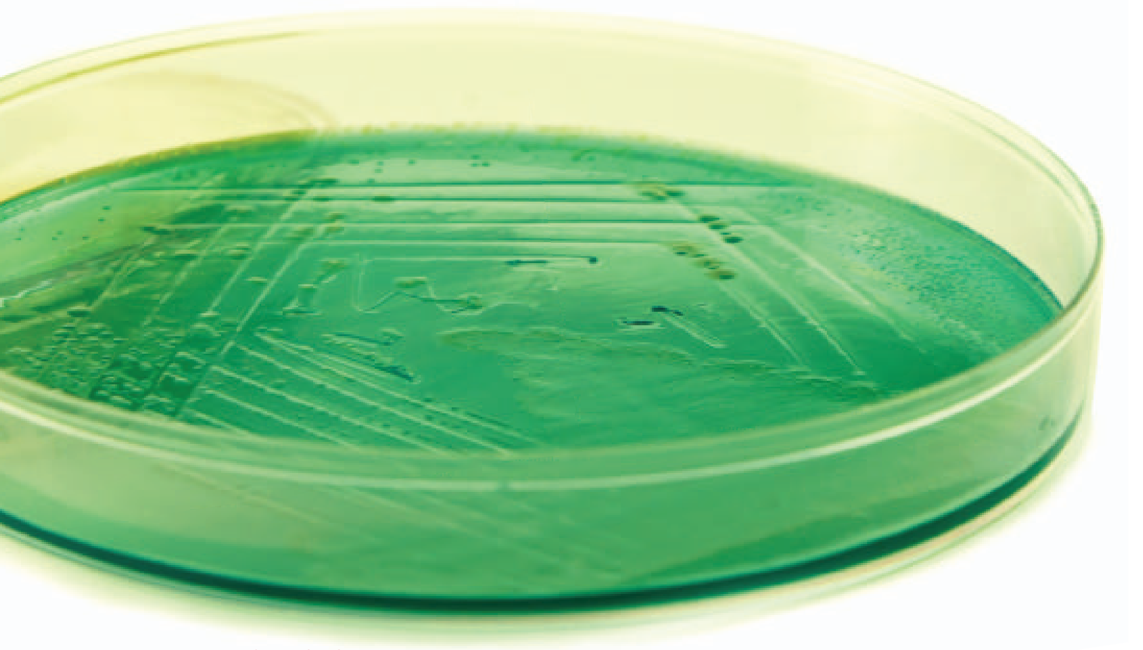
**Full text:** dx.doi.org/10.1126/science.abi9756

Where Science Gets Social.
AAAS.ORG/COMMUNITY
Share
Member COMMUNITY
AAAS
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

# RESEARCH

## IN *SCIENCE* JOURNALS

*Edited by* **Michael Funk**



NATURAL PRODUCTS

### A copper-containing antibiotic

Bacteria require transition metal ions for biological processes and must also protect themselves against excess metal, which is toxic. Patteson *et al.* explored how the environmental bacterium *Pseudomonas aeruginosa* uses a five-enzyme pathway to synthesize a small-molecule complex, fluopsin C, which is built from cysteine and contains a copper ion. The biosynthesis involves unusual enzymatic transformations that convert cysteine to a thiohydroximate, two of which chelate a copper ion in the final natural product. Fluopsin C protects *P. aeruginosa* from excess copper and also acts as a broad-spectrum antibiotic against other bacteria. —VV *Science*, abj6749, this issue p. 1005

**The ubiquitous bacterium *Pseudomonas aeruginosa* is a prolific producer of natural products, including a copper-containing antibiotic.**

INFECTIOUS DISEASES

### Messenger RNA vaccine against ticks

Repeated exposures to the black-legged tick, *Ixodes scapularis*, can lead to acquired resistance against ticks. To bypass the need for repeated exposures, Sajid *et al.* developed a messenger RNA vaccine that encoded for 19 *I. scapularis* salivary proteins. Vaccinated hamsters developed erythema at the site of tick attachment, a feature of acquired tick resistance. The immunization also led to poor tick feeding and reduced transmission of a tick-borne bacterial pathogen. —CSM

*Sci. Transl. Med.* **13**, eabj9827 (2021).

DATA SCIENCE

### Privacy risks of public genomic data

To promote biomedical research, genomic data are often shared in controlled settings or public repositories, posing a privacy risk if they can be algorithmically re-associated with an individual. Venkatesaramani *et al.* show that the average risk to privacy in this setting is relatively low and decreases further with more individuals in the data. The authors analyzed public genomes and face images to link faces to phenotypes (e.g., eye, hair, and skin color) using current deep learning tools and to link genomes to phenotypes using single-nucleotide polymorphisms. However, technological advances could pose new risks, and targeted privacy risks remain. These findings highlight the potential for future privacy risks as technology continues to advance. —AJC

*Sci. Adv.* 10.1126/sciadv.abg3296 (2021).

CORONAVIRUS

### Deadly surge in Delhi

In the spring of 2021, Delhi, India, experienced a wave of coronavirus cases that overwhelmed health care services despite the population showing a high level of immune positivity. To study this, Dhar *et al*. collated a mixture of serosurveillance, quantitative polymerase chain reaction, and genomic data, finding that waves of variants had passed through the Delhi population during 2020 and 2021. The Alpha (B.1.1.7) variant dominated in March 2021 and was rapidly replaced by the Delta (B.1.617.2) variant in April and May 2021. The Delta variant outcompeted its predecessors by mutations that enhanced replication, immune evasion, and host receptor avidity, thus increasing transmissibility, reinfection, and vaccination breakthrough. —CA

*Science*, abj9932, this issue p. 995

VIROLOGY

### A tricky mimicry

RNA viruses use dynamic, multifunctional folded elements to hijack host cellular machinery. Bonilla *et al*. used cryo–electron microscopy (cryo-EM) to explore an RNA element from Brome mosaic virus that tricks host cell tyrosine transfer RNA synthetase (TyrRS) into adding a tyrosine to the viral genome's 3′ end. Visualizing this RNA both in isolation and bound to a cellular TyrRS revealed a bound structure unlike the canonical transfer RNA–like L-shaped and conformational rearrangements in the RNA upon binding to the TyrRS,

PHOTO: ZAHARIA BOGDAN RARES/SHUTTERSTOCK

which suggests a multistep process of enzyme recognition. This study highlights the power of cryo-EM to illustrate dynamic processes involving small structured RNAs and RNA-protein complexes. —DJ

*Science*, abe8526, this issue p. 955

QUANTUM GASES

## Photons not welcome

Two identical fermions cannot occupy the same quantum state, or so says the Pauli principle. For a cold gas of fermionic atoms, this means that all states up to the Fermi energy will be occupied, with only the atoms with the highest energy able to change their state. Such conditions have long been predicted to suppress light scattering off gases because the atoms receiving a kick from collisions with photons would have no state to move to. Deb *et al.*, Margalit *et al.*, and Sanner *et al*. now describe this so-called Pauli blocking of light scattering. —JS

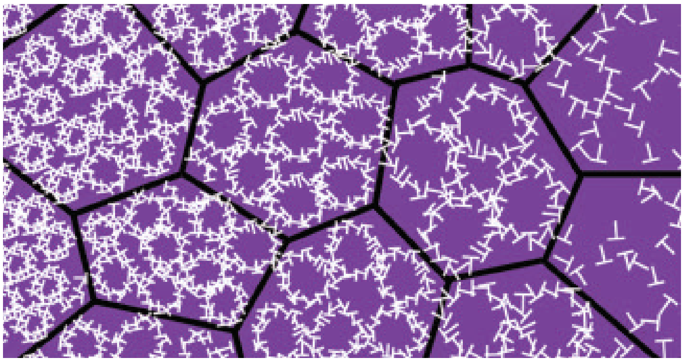
*Science*, abh3470, abi6153, abh3483, this issue p. 972, this issue p. 976, this issue p. 979

METALLURGY

## Strength by cyclic torsion

For most alloys, conventional or high entropy, increasing strength comes at the cost of poor ductility. Although there are many strategies to break this inverse relationship, Pan *et al*. show that cyclic torsion on a high-entropy alloy enhances strength without degrading ductility (see the Perspective by Yeh). Cyclic torsion creates a gradient of dislocations and low-angle grain boundaries from the surface to the interior that organize into tiny stacking faults and twin when straining begins. These structures allow for good ductility while simultaneously helping to work harden the alloy. —BG

*Science*, abj8114, this issue p. 984; see also abm0120, p. 940



Cyclic torsion generates a gradient of dislocations that provides both ductility and hardness to multicomponent high-entropy alloys.

CORONAVIRUS

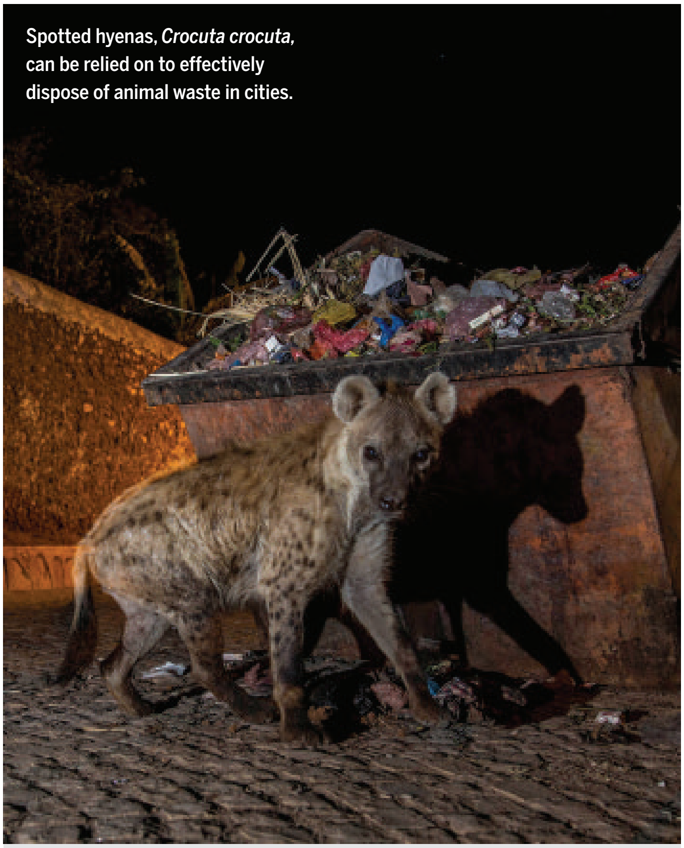
## SARS-CoV-2: To have or to have not

In June 2021, official records in Kenya showed fewer than 4000 confirmed deaths and 180,000 confirmed cases of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). These data tend to reflect the economically advantaged strata of society who can afford smartphones and have access to medical attention and tests. Brand *et al*. developed an epidemiological model to estimate the impact of the pandemic in Kenya, the population of which was split into two socioeconomic strata. The authors predicted that 75% of the Kenyan population (about 39 million people) had been exposed to the virus by June 2021. If a fourth wave of infection is observed in the future, it would likely be driven by a variant with enhanced transmissibility or natural immune escape. —CA

*Science*, abk0414, this issue p. 989

# IN OTHER JOURNALS

*Edited by* **Caroline Ash** *and* **Jesse Smith**



Spotted hyenas, *Crocuta crocuta*, can be relied on to effectively dispose of animal waste in cities.

ECOSYSTEM SERVICES

## Clean up brigade

Hyenas are considered unattractive and feared by some, but they are charismatic in unexpected ways, not least because of their vocalizations but also because of their liking for discarded human waste. Sonawane *et al.* investigated the contribution that hyenas make to waste cleanup in the Ethiopian city of Mekelle. This city has inadequate organized waste collection, causing a significant public health risk. Over 40 nights, Mekelle's 200 or so hyenas consumed all of the animal waste available in the city's landfills. This equated to about a metric ton per year per hyena. These activities were estimated to reduce anthrax and bovine tuberculosis transmission to humans and livestock by about 4%, saving approximately US $52,000. Rather than vilifying hyenas, we should welcome their attention and optimize their scavenging opportunities. —CA

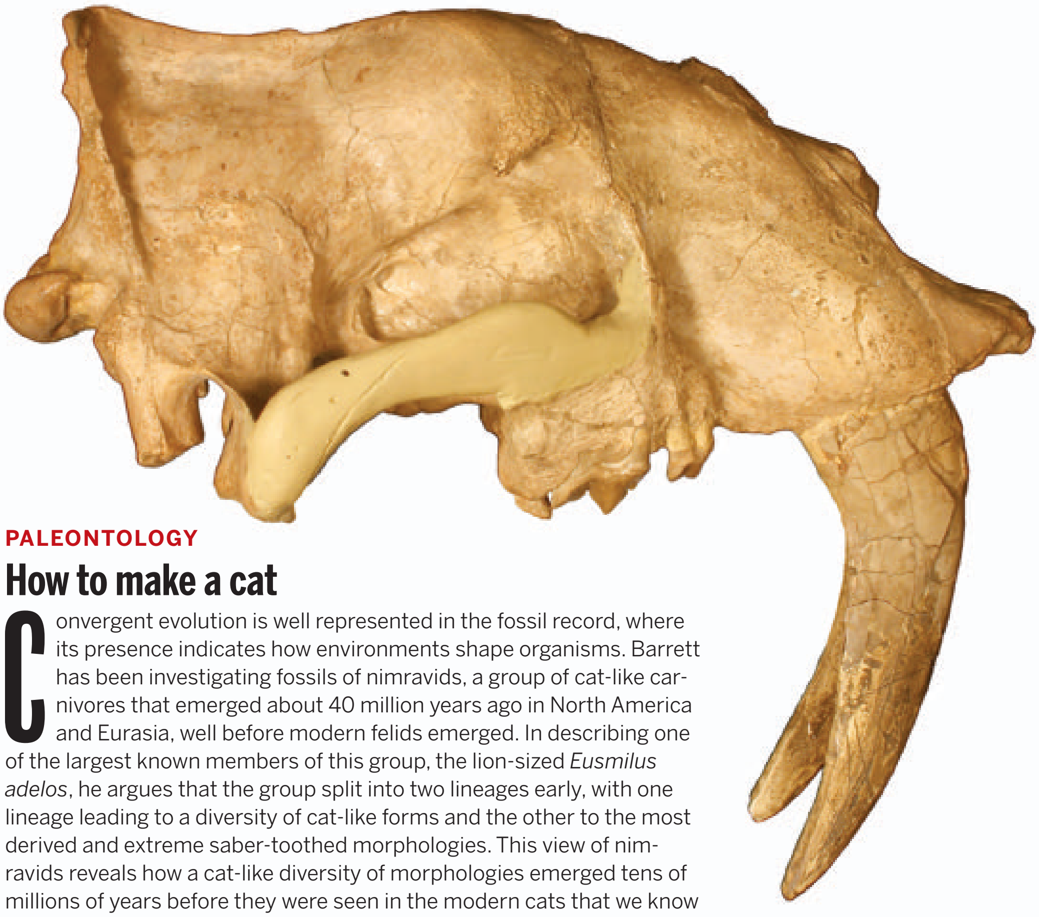
*J. Anim. Ecol.* 10.1111/1365-2664.14024 (2021).

CATALYSIS

## Pore-dependent water structures

The confined space in zeolite pores and interactions through hydrogen bonding with silanol OH groups can create solvent structures that can be more reactive than bulk water. Bregante *et al.* examined the effect of pore size on water structure using both infrared spectroscopy and molecular dynamics simulations. For the large 1.3-nanometer pores of

CREDITS: (LEFT TO RIGHT) IMAGE: PAN *ET AL.*; LUKE MASSEY/MINDEN PICTURES



PALEONTOLOGY

## How to make a cat

Convergent evolution is well represented in the fossil record, where its presence indicates how environments shape organisms. Barrett has been investigating fossils of nimravids, a group of cat-like carnivores that emerged about 40 million years ago in North America and Eurasia, well before modern felids emerged. In describing one of the largest known members of this group, the lion-sized *Eusmilus adelos*, he argues that the group split into two lineages early, with one lineage leading to a diversity of cat-like forms and the other to the most derived and extreme saber-toothed morphologies. This view of nimravids reveals how a cat-like diversity of morphologies emerged tens of millions of years before they were seen in the modern cats that we know today. —SNV *Sci. Rep.* **11**, 21078 (2021).

**Lateral view of a fossil skull of the lion-size nimravid *Eusimilus adelos*, a cat-like lineage that evolved millions of years before modern cats**

faujasite, the water structure resembled that of bulk, but for pore sizes below 0.65 nanometers in other zeolites, oligomeric one-dimensional chains formed. The enthalpic penalty for the reorganization of the structures during aqueous catalysis was more than offset by entropic gains, which was reflected in higher rates of hexene epoxidation (up to 400-fold) in smaller pores. —PDS

*Nat. Catal.* **4**, 797 (2021).

HOST DEFENSE

## Leishmaniasis? NO means no!

Nitric oxide (NO) helps to limit the tissue damage caused by infections. It can do so directly, through its direct antimicrobial effects, or indirectly, by inhibiting the recruitment of inflammatory cells such as monocytes and neutrophils. Formaglio *et al.* found that inhibition of inflammation is crucial to controlling the intracellular parasite *Leishmania major*. Combining mathematical modeling with intravital imaging of infected mice, the authors show that NO dampens *L. major* proliferation by depriving access to monocytes, one of its preferred niches. Direct killing by NO is important only during the peak of infection. It remains to be seen whether NO similarly suppresses other protozoan pathogens, such as *Trypanosoma cruzi*, which preferentially infect phagocytes. —STS

*Immunity* 10.1016/j.immuni.2021.09.021 (2021).

CELL BIOLOGY

## Here we grow

A long-standing controversy surrounds the occurrence of local translation in axons, partly because of the inability to visualize ribosomes directly in these neuronal extensions. To gain visual access to actively growing human axons, Hoffmann *et al.* combined cryo–correlative fluorescence microscopy and electron tomography with human cerebral organoid technology. This approach allows imaging of the subcellular architecture of axonal shafts at very high resolution in close to in situ conditions. The structure of the endoplasmic reticulum in growing axons suggests a primary function for local production of membrane lipids. By contrast, the relative lack of ribosomes in axon shafts indicates that proteins are produced in the bulk of the cytosol. This technology offers a promising tool for probing the behavior of many kinds of specialist cells. —SMH

*eLife* **10**, e70269 (2021).

BIOMEDICINE

## Sensing sutures

Many complications can arise after surgery, including infection, bleeding, or reopening of the wound. When the wound site is in deep tissue, these problems are often only detected by a change in coloration or body temperature, at which point complications may be severe. Kalidasan *et al.* developed multi-filament surgical sutures with imbedded absorbent sensors. The sutures can be operated wirelessly using a radiofrequency identification technique to monitor wound integrity, leakage of gastric fluid, and tissue micromotions, as shown in deep surgical sites in a porcine model. Two-week monitoring regimens in rats showed that the sutures produced similar healing outcomes as medical grade ones. —MSL

*Nat. Biomed. Eng.* **5**, 1217 (2021).

POLLUTION MITIGATION

## The long and short games

Mitigating anthropogenic climate change may have immediate costs, but the long-term benefits of pollution reduction are greater—even greater than previous estimates—says a study by Shindell *et al*. The authors show how emission reductions affect not just the trajectory of climate change, with its consequent reduction of climate-related costs after several decades, but also how the immediate improvement of air quality results in lower costs in the areas of human health, labor productivity, and crop yield within the first 10 years of implementation. Therefore, reducing emissions of carbon dioxide and other pollutants is a winning strategy in both the long and short terms. —HJS

*Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104061118 (2021).

IMAGE: BARRETT, *SCI. REP.* **11**, 21078 (2021).

# ALSO IN *SCIENCE* JOURNALS

*Edited by* **Michael Funk**

DEVELOPMENTAL BIOLOGY

## X chromosome activity in primates

Mammalian X and Y sex chromosomes differ greatly in structure and gene content. Therefore, mechanisms are in place to ensure that there is balanced gene dosage between the X chromosome and autosomes and between the sexes. Quite a bit is known about dosage compensation in mice; however, the mechanism in primates, including humans, is less clear. Okamoto *et al.* show in cynomolgus monkeys that the X chromosomes display repressive modifications and compacted structure before being chosen as active or inactive X chromosomes (see the Perspective by Heard and Rougeulle). This work elucidates the pathway, timing, and progression of X chromosome dosage compensation during monkey development, creating a foundation for elucidating this mechanism in other primates, including humans. —BAP

*Science*, abd8887, this issue p. 954; see also abm1857, p. 942

QUANTUM MECHANICS

## Double slits with molecular states

Despite decades of research, the role of quantum mechanical effects in molecular scattering has not yet been fully investigated and can still demonstrate fascinating results, even for simple triatomic systems. Zhou *et al.* show that the entangled bond axis orientations in the biaxial state of a deuterium molecule can act as the two slits of a double-slit interferometer for rotationally inelastic collision with a helium atom, giving rise to quantum interference between two indistinguishable pathways (see the Perspective by Wang and Yang). The present work presents an elegant example of quantum interference in molecular scattering that is conceptually similar to the famous Young's optical double-slit experiment. The proposed molecular interferometer could be used to coherently control the phases in various molecular processes in future experiments. —YS

*Science*, abl4143, this issue p. 960; see also abm5536, p. 938

JUPITER

## Measuring the depth of Jupiter's storms

The atmosphere of Jupiter consists of bands of winds rotating at different rates, punctuated by giant storms. The largest storm is the Great Red Spot (GRS), which has persisted for more than a century. It has been unclear whether the storms are confined to a thin layer near the top of the atmosphere or if they extend deep into the planet. Bolton *et al.* used microwave observations from the Juno spacecraft to observe several storms and vortices. They found that the storms extended below the depths at which water and ammonia are expected to condense, implying a connection with the deep atmosphere. Parisi *et al.* analyzed gravity measurements taken while Juno flew over the GRS. They detected a perturbation in the planet's gravitational field caused by the storm, finding that it was no more than 500 kilometers deep. In combination, these results constrain how Jupiter's meteorology links to its deep interior. —KTS

*Science*, abf1396, abf1015, this issue p. 964, this issue p. 968

DENGUE EVOLUTION

## Variations in disease enhancement

Secondary Dengue virus (DENV) infections can be dangerous if levels of antibodies from prior infection are inadequate to clear the virus. This RNA flavivirus exploits the presence of lower levels of heterotypic antibodies to infect immunoglobulin Fcγ receptor–bearing cells. Many RNA viruses also exhibit antigenic variation, which classically allows evasion of immune responses. Katzelnick *et al.* investigated whether antigenic variation in DENV has a biological function in a virus that courts immune responses to enhance replication (see the Perspective by Rohani and Drake). Using antigenic cartography on a panel of more than 400 DENV1-4 subtype samples isolated in Bangkok, Thailand, the authors found that antigenic variation in virus populations oscillated between similarity and dissimilarity across subtypes over time, with outbreaks correlating with periods of antigenic dissimilarity within serotypes. This pattern may be at least in part a result of the conflicting evolutionary pressures of immune evasion and immune enhancement. —CA

*Science*, abk0058, this issue p. 999; see also abm6812, p. 941

NEUROSCIENCE

## How the body regulates fear

Although fear is important for survival, it is maladaptive if it is either too strong, as in anxiety disorders, or too weak, as in exaggerated risk taking. Working in mice, Klein *et al.* observed that the insular cortex has an unparalleled dual role in either enhancing or weakening the extinction of fear, depending on the internal fear state of the animal (see the Perspective by Christianson). This insula function helps to maintain fear within a homeostatic range and depends on bodily feedback signals: Fear-induced freezing behavior is associated with a slowed heart rate, which in turn dampens fear-evoked activity of the insular cortex. Two opposite signals, prediction of threat by fear-associated cues and negative feedback signals from the body, are thus integrated within the insular cortex. —PRS

*Science*, abj8817, this issue p. 1010; see also abm6790, p. 937

PHYSIOLOGY

## Fattening livers with thyroid hormone

The thyroid hormone T3 can promote fatty acid oxidation in the liver, but it can also have the opposite effect on hepatic lipid metabolism. Mendoza *et al.* investigated this dichotomy in mice with hepatocyte-specific genetic deficiencies on diets that mimicked hypothyroidism or hyperthyroidism. T3 enhanced de novo lipogenesis in the liver through the transcription factor carbohydrate response element binding protein (ChREBP), which was recruited to the promoters of lipogenic genes by the T3 receptor TRβ1. Mice with a hepatocyte-specific deficiency of ChREBP had decreased hepatic triglyceride amounts in response to T3. —WW

*Sci. Signal.* **14**, eabh3839 (2021).

AGING

## Caution around the fountain of youth

The scientific and popular literature is full of claims for diets that delay or reverse the aging process (at least in model organisms). But how do these interventions work? Is it the amount of food, the timing of food intake, or the proportion of certain macronutrients? In a Review, Lee *et al.* explore the fact and fiction of dietary prescriptions for a healthier and longer life. They propose that one unifying concept may be convergence on the signaling pathway mediated by the protein kinase mTOR (mechanistic target of rapamycin). Another conclusion is that the efficacy and safety of these diets for humans largely remain to be established. —LBR

*Science*, abe7365, this issue p. 953.

TOLERANCE

## Two-faced inducers of peripheral tolerance

Autoimmune regulator (*Aire*), which enables the expression of diverse self-antigens in medullary thymic epithelial cells (mTECs), is also found in extrathymic Aire-expressing cells (eTACs), but their precise identity and biology has remained controversial. Wang *et al.* used single-cell multiomics and transgenic Aire-reporter mice to clearly define these populations, revealing that eTACs consist of two related populations that have transcriptional homology with mTECs: Aire-expressing migratory dendritic cells and a unique subset of this group designated Janus cells, which also express a receptor called ROR$\gamma$t. Self-antigen expression in eTACs is sufficient to delete T cells escaping thymic selection, preventing autoimmune diabetes. These findings establish eTACs as subsets of dendritic cells equipped to sustain tolerance beyond the thymus and reveal the symmetry of a core Aire-associated transcriptional network. —IRW

*Sci. Immunol.* **6**, eabl5053 (2021)

REVIEW SUMMARY

AGING

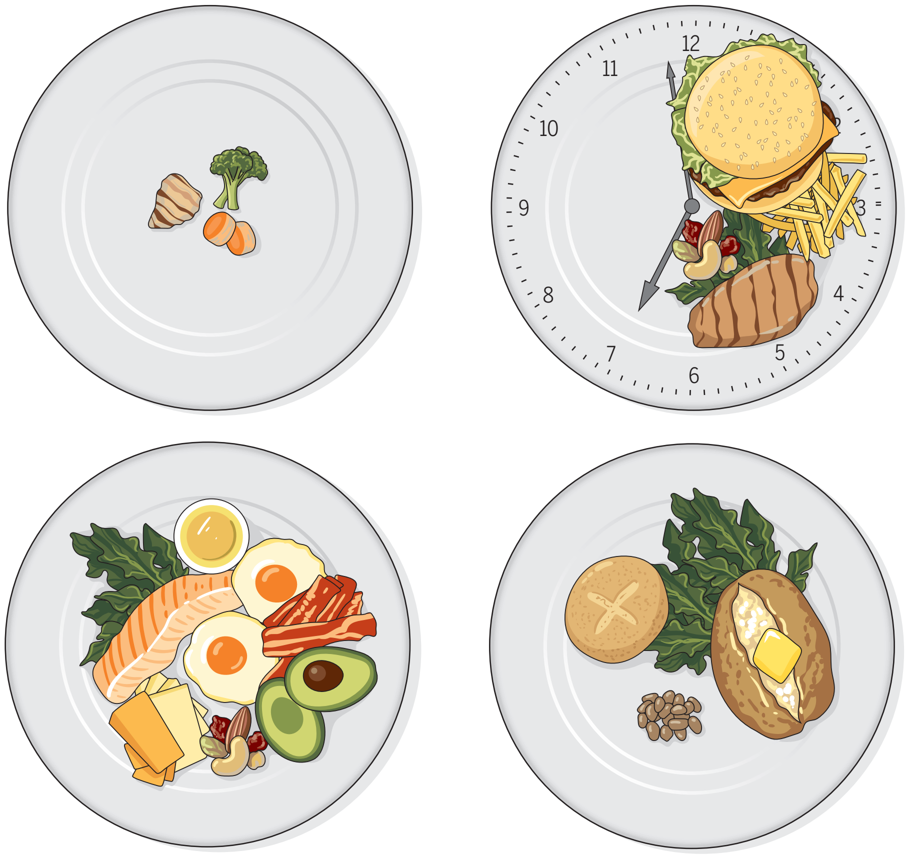
# Antiaging diets: Separating fact from fiction

Mitchell B. Lee, Cristal M. Hill, Alessandro Bitto, Matt Kaeberlein*

**BACKGROUND:** Reduced caloric intake without malnutrition is the oldest known life span–extending intervention. Laboratory studies throughout the 20th century established and confirmed the benefits of caloric restriction (CR) in multiple model systems. CR not only increased life span across evolutionarily distant organisms but also reduced age-associated disease burden and functional decline in these studies. Epidemiological data from human populations is also generally consistent with the idea that lower caloric intake is associated with increased life expectancy. In recent years, numerous diet modalities that are purported to be "antiaging" have sprung from these observations. These diets restrict particular macronutrients (carbohydrates or protein) or feeding intervals and can be divided into those that impose reduced caloric intake versus those that are isocaloric to control diets.

**ADVANCES:** We evaluated several of the most popular antiaging diets, including CR, intermittent fasting, fasting-mimicking diets, ketogenic diets, time-restricted feeding, protein restriction, and essential amino acid restriction. By characterizing these nutritional interventions in comparison with classical CR, we gained numerous insights. Many studies fail to control for reduced caloric intake in the diet group, making their effects impossible to decouple from CR. Although often presented as uniformly beneficial, the effects of CR on life span are highly dependent on genotype and, in some cases, cause reduced survival. Despite their limitations, these studies have greatly improved our understanding of the cellular response to low nutrient availability. A picture is beginning to emerge of a complex network composed of multiple signaling pathways that converge on key molecular hubs; foremost among these is the mechanistic target of rapamycin (mTOR). Because mTOR and other components of this network are well-studied drug targets, there continues to be considerable interest in pharmacologically targeting this network to increase longevity and health span. Human studies, both correlative and controlled, are consistent with health benefits conferred by a CR diet. However, it remains unresolved whether these benefits are a consequence of modulating the aging process itself or are simply the result of avoiding obesity. Several unresolved questions suggest caution when considering whether to recommend or implement any of these diets among the healthy general public. Among these is understanding how genetic and environmental variation modify diet response, especially in understudied populations and in the context of environmental challenges such as, for example, a global viral pandemic.

**OUTLOOK:** CR and other antiaging diets have yielded important insights into the complex and evolutionarily conserved signaling pathways that transduce information regarding environmental nutrient availability into a physiological response to promote healthy longevity. This understanding, in turn, has opened the door to a new generation of longevity-promoting interventions that mimic molecular responses to nutrient deprivation. Although CR and other diets hold promise, additional data from carefully controlled studies is needed before broadly recommending or implementing these diets, or other interventions, for otherwise healthy people. Human genetic and environmental variation combined with the challenge of modeling human aging in ultimately dissimilar mammalian model systems pose fundamental limitations to our current ability to predictably translate these findings to people. From a pragmatic perspective, even if these challenges can be overcome, widespread adoption of dietary interventions for healthy longevity seem unrealistic. We therefore suggest that alternative, nondietary strategies with the potential for public uptake should therefore be pursued. In particular, validated biomarkers of biological aging are required to match intervention to each person's distinct genetic and environmental context and thereby optimize individual healthy life span. Future research directed at clarifying the underlying mechanisms involved in eliciting the longevity-promoting response to CR, and how this differs among individuals, should one day help us realize a true precision geroscience approach. ■



**At the buffet of antiaging diets, which is the best plate?** Diets clockwise from top left: CR, time-restricted feeding, protein restriction, and ketogenic.

IMAGE: KELLIE HOLOSKIE/SCIENCE

---

The list of author affiliations is available in the full article online.

*Corresponding author. Email: kaeber@uw.edu

Cite this article as M. B. Lee *et al.*, *Science* **374**, eabe7365 (2021). DOI: 10.1126/science.abe7365

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abe7365

REVIEW

AGING

# Antiaging diets: Separating fact from fiction

**Mitchell B. Lee[1], Cristal M. Hill[2], Alessandro Bitto[1], Matt Kaeberlein[1]***

**Caloric restriction has been known for nearly a century to extend life span and delay age-associated pathology in laboratory animals. More recently, alternative "antiaging" diet modalities have been described that provide new mechanistic insights and potential clinical applications. These include intermittent fasting, fasting-mimicking diets, ketogenic diets, time-restricted feeding, protein restriction, and dietary restriction of specific amino acids. Despite mainstream popularization of some of these diets, many questions remain about their efficacy outside of a laboratory setting. Studies of these interventions support at least partially overlapping mechanisms of action and provide insights into what appear to be highly conserved mechanisms of biological aging.**

Modern aging research can trace its roots back to studies from the early 1900s that examined the effects of reduced food intake on life span in rats (*1*, *2*). These pioneering experiments showed that reducing caloric intake in laboratory-reared animals delays development and results in a substantial increase in adult life span. Work by Weindruch, Masoro, Walford, and others in the 1970s, 1980s, and 1990s expanded and popularized this area of research in both rats and mice and established caloric restriction (CR) as the dominant paradigm for antiaging intervention (Box 1) for the rest of the 20th century (*3*). These foundational studies provided strong evidence that CR not only increases life span in rodents but also reduces disease burden and delays many functional declines of old age (*4*).

A common definition for CR in these early studies is "reduced caloric intake in the absence of malnutrition" (*5*). Typically, this was accomplished by limiting chow by a fixed amount while supplementing with vitamins and micronutrients. The precise method of food limitation varied (for example, intermittent versus continuous feeding), as did the timing of initiation (pre- or post-weaning) and degree of restriction. The data from these studies support the idea that at least in some common laboratory strains of both mice and rats, total caloric intake correlates inversely with life span up to about 50 to 60% restriction (so long as essential nutrition is maintained), and starting earlier in life gives larger effects on life span than starting later in life (*5*).

With popularization of invertebrate models in aging research near the end of the 20th century, it was natural for scientists to test whether CR could similarly affect aging in these organisms. Because culture conditions differ across species, multiple alternative methods of nutritional intervention were tested and found to increase life span. We will refer to these interventions collectively using the term "dietary restriction" (DR), which includes both sugar and amino acid limitation in budding yeast, reduced bacterial food availability in nematode worms, and lower levels of protein (yeast extract) or sugar in fruit flies (*6*). These foundational studies allowed for mechanistic analyses not previously feasible in rodents and identified a highly conserved nutrient-sensing, growth-promoting network that appears to regulate biological aging in many different organisms. Among the key proteins in this network are the mechanistic target of rapamycin (mTOR), adenosine 5′-monophosphate (AMP)–activated protein kinase (AMPK), insulin and insulin-like growth factor 1 (IGF-1)–like receptors, FOXO-family transcription factors, and nicotinamide adenine dinucleotide (NAD)–dependent sirtuin deacetylases (*7*, *8*).

As the molecular underpinnings of CR, and DR more generally, became established, attention shifted toward identifying small molecules that mirror the effects of CR on life span and health without requiring reduced food consumption. Such "CR mimetics" include the mTOR inhibitor rapamycin, the antidiabetes drug metformin, the glycolytic inhibitor 2-deoxyglucose, the intestinal $\alpha$-glucosidase inhibitor acarbose, and sirtuin-activating compounds (*9*, *10*). Most putative CR mimetic compounds, with the possible exception of rapamycin, have thus far failed to match the magnitude of life span extension and health span benefits of CR. For example, metformin (a nonspecific AMPK activator) and resveratrol (a nonspecific sirtuin activator) primarily improve measures of metabolic health during aging in mice, including increased insulin sensitivity and reduced cancer incidence, without reproducibly extending life span (*11*, *12*). These discrepancies likely reflect a still incomplete understanding of the varied and diverse effects of CR, which have yet to be fully recapitulated pharmacologically.

## The rise of antiaging diets

The goal of this critical Review is to summarize the current state of the field with respect to the most commonly studied antiaging dietary interventions, with a focus on potential shared mechanisms of action, important unanswered questions and areas for future inquiry, and addressing common misconceptions in the literature. We largely restrict our considerations to preclinical studies in rodents and, where applicable, relevant human data. For detailed reviews of CR in rodents, we refer the interested reader to these resources (*4*, *5*, *13*–*15*). Here, we only briefly discuss the evidence for antiaging effects of classic CR and instead focus on placing alternative dietary interventions into context with what is known about CR and potential impacts on human health and longevity. Among the interventions we consider are ketogenic diets (KDs), intermittent fasting (IF), fasting-mimicking diets (FMDs), time-restricted feeding

### Box 1. Reclaiming the term "antiaging."

The phrase "antiaging" is greatly abused in popular culture, often for the purpose of marketing cosmetic procedures or unproven nutritional supplements purported to slow or reverse aging. This has the unfortunate consequence of creating confusion among the general public and diminishing the impact of legitimate scientific discovery. Here, we define "antiaging" as delaying or reversing biological aging by targeting the established molecular mechanisms of aging, which have been formalized as "hallmarks" or "pillars" of aging (*93*, *94*). Effective antiaging interventions in laboratory animals increase both median and maximum population life span and broadly delay the onset and progression of many age-related functional declines and diseases. The latter effect is often referred to as "extending health span," which is a qualitative term referring to the period of life free from chronic disease and disability (*95*). Recent studies show that at least some antiaging interventions, such as the drug rapamycin, can reverse functional declines across multiple tissues in aged animals (*96*). On the basis of this definition, there are as yet no clinically validated antiaging interventions in humans. However, there is some evidence consistent with antiaging effects for CR and related diets in humans as well as a small number of putative geroprotective compounds, including metformin and rapamycin (*97*).

---

[1]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA 98195-7470, USA.
[2]Pennington Biomedical Research Center, Baton Rouge, LA 70808, USA.
*Corresponding author. Email: kaeber@uw.edu

(TRF), protein restriction (PR), and diets restricted for specific amino acids, including methionine restriction, tryptophan restriction, and branched chain amino acid (BCAA) restriction.

One critical but often overlooked (and therefore quite confusing) consideration when evaluating different antiaging diets is the relative caloric intake of control and experimental cohorts. Antiaging diets can be considered in two broad groups: CR and isocaloric nutrient restriction (Table 1). Several of the most prominent antiaging diets such as IF, FMDs, and KDs generally fall under the CR umbrella because the experimental group typically consumes 20 to 40% fewer calories than does the control group. This makes evaluating the effects of dietary composition challenging to differentiate from the effects of reduced caloric intake. Other interventions, such as TRF and PR or amino acid restriction are somewhat better characterized under isocaloric experimental conditions. In all cases, however, one should carefully evaluate relative caloric intake when considering studies in this area.

### *KDs*

KDs refer to dietary compositions designed to maintain a constant state of ketogenesis, the metabolic production of ketone bodies (acetoacetate, β-hydroxybutyrate, and acetone) as a by-product of fat metabolism in the liver. This results in ketosis, a state of elevated ketone bodies in the blood that can then be taken up and metabolized by other tissues. KDs have been studied in humans for many decades as a treatment for epilepsy and have achieved mainstream popularity because of their palatability and effectiveness at inducing weight loss (*16*). In humans, the most common KD is typically very low in carbohydrates (less than 30 to 50 g per day), with ~75% of calories derived from fats (*17*). Many other KD variations are possible, so long as carbohydrate levels remain low enough to induce ketogenesis, such as the popular high-protein Atkins Diet (*18*). Currently, the long-term health consequences of KDs in humans and the relative merits of low- versus high-protein KD diets are vigorously debated within the nutrition community, with little consensus beyond clear efficacy for epilepsy and weight loss.

KDs recently gained recognition for potential effects on biological aging with two 2017 papers that reported that a low-carbohydrate, low-protein KD is sufficient to increase mean life span as well as health span measures in mice (*19*, *20*). In one study, a 0%-carbohydrate KD that achieved high levels of β-hydroxybutyrate in blood was initiated at 12 months of age and given to the mice either continuously or in a cyclic fashion interspersed with control chow on a weekly basis (*19*). The continuous KD failed to increase life span, whereas the cyclic KD increased mean, but not maximum, life span and improved both memory function and metabolic parameters late in life. In the other study, a low-carbohydrate diet (12% carbohydrates) or a KD (less than 1% carbohydrates) were initiated at 12 months of age (*20*). Both diets appeared to increase median life span relative to control-fed animals, with the KD resulting in a 13% increase in median life span and trend toward a smaller increase in maximum life span, which did not reach statistical significance. Improvements in memory, motor function, and reduced cancer incidence were observed in the mice fed a KD in this study. Both studies observed reduced mTOR activity in the longer-lived mice eating a KD.

One question is whether KD effects are mediated by ketone bodies directly. Ketone bodies produced by the liver enter circulation and are taken up by other tissues, where they can directly enter the tricarboxylic acid cycle through acetyl–coenzyme A, bypassing the need for glycolytic breakdown of glucose. The ketone body β-hydroxybutyrate has also been implicated as a signaling molecule that can act through extracellular receptors to regulate histone acetylation and thereby impact gene expression (*21*). One study reported that β-hydroxybutyrate supplementation could extend life span in *Caenorhabditis elegans* through a mechanism linked to reduced mTOR signaling (*22*), although there is not yet direct evidence

**Table 1. Summary of antiaging diets.** Life span effect refers to studies in rodents. Number of arrows are intended to indicate relative robustness and consistency of reported effects.

| Dietary intervention | Description | Life span effect |
|---|---|---|
| **Low-calorie interventions** | | |
| "Classic" CR | Daily reduction in calories, typically by 20 to 50%, without malnutrition. Macronutrient ratios are unchanged. | ↑↑↑ |
| CR without PR | CR in which protein content is modified so that only calories are reduced and protein intake is not changed. | ↑↑↑ |
| IF | CR variant with at least 1 day of fasting between feedings. Many classic CR studies used intermittent fasting protocols in which mice were fed three times per week. | ↑↑↑ |
| FMD | Cyclic CR in which a low-calorie KD is provided during the restricted phase. In mice, FMD cycles are typically 3 to 4 days followed by 3 days of refeeding. | ↑↑ |
| KD | Restriction of carbohydrates to induce ketosis. In mice, carbohydrates are limited to less than 1% of total calories. KDs do not have to be low calorie, but the variations studies in mice resulted in reduced caloric consumption. | ↑ |
| **Isocaloric diets** | | |
| PR | In mice and rats, isocaloric PR has been reported to extend life span, but the effects appear to be much smaller than CR and may be sex-specific in mice.* | ↑ |
| Essential amino acid restriction | Restriction of methionine, tryptophan, or BCAA content in the diet. Essential amino acid restriction in mice typically involves reducing methionine by about 80%, tryptophan by about 40%, or BCAAs by about 67%. It remains unclear what extent these interventions share similar mechanisms. | ↑ |
| TRF | Ad libitum feeding restricted to a specific period of the day. In people, a common TRF protocol is 16:8 (hours fasting: hours feeding). In mice 12:12 has been tested.† | ↑ |
| Isocaloric IF | IF where the IF group consumes an equal number of calories as the control group by overfeeding during the *ad libitum* phase.‡ | ↑ |

*A recent report found that PR increased life span in male mice but not female mice (*55*).

†One study of only male mice reported 11% life span extension in "isocaloric" TRF mice, but the experimental mice consumed slightly less kcal/day than did the control mice (*40*).

‡Every-other-day feeding increased life span in one study by 13% under roughly isocaloric conditions (*92*).

that ketone body metabolism is sufficient to confer increased life span in mice. Ketone esters are reported to reduce anxiety-like behaviors and decrease amyloid-β and amyloid-τ deposits in a mouse Alzheimer's disease model (*23*) and have also been reported to reduce blood insulin and glucose levels and to inhibit mTOR signaling (*24*). Taken together, these findings are highly suggestive that ketone esters themselves could have antiaging properties and highlight the importance of additional research in this area.

### *IF and FMDs*

Fasting has long been touted for its putative health benefits in different cultures, and recent years have seen a resurgence of research on possible antiaging effects of diets that incorporate fasting or "fasting-mimicking" components. Reviews of this topic will often make blanket statements about the health benefits of IF for numerous age-related conditions in mice, and although there is certainly evidence to support such assertions, the experimental protocols used generally amount to CR (*25*, *26*); it appears that modern IF protocols are largely a rebranding of classic CR methods [for example, (*3*), where the CR mice were fed three times per week]. Hungry laboratory mice (and any CR researcher will tell you that their mice are hungry) will typically eat all of the available food in their cage quickly compared with control-fed mice. Thus, even in a CR study in which mice are fed daily, the restricted mice will typically be fasting for at least 18 hours between meals. This is not intended to downplay the potential importance of the physiological changes associated with fasting, such as ketogenesis, but is simply an observation that the majority of preclinical studies referring to fasting-based experimental protocols cannot be differentiated from CR because researchers rarely ensure isocaloric conditions (Table 1).

Over the past decade, FMDs have emerged as a highly studied cyclic variant of CR designed to induce metabolic responses akin to fasting through a nutrient-dense, low-calorie diet. FMDs induce ketogenesis by restricting protein and simple carbohydrates while maintaining high fat levels (*27*, *28*) and could therefore be considered as an intermittent KD. Initial studies in rodents showed that bimonthly 4-day FMD reduces both body and organ size and improves a wide range of age-related parameters, including adiposity, tumor burden, motor and cognitive function, neurogenesis, and median (but not maximum) life span (*27*). One interesting effect of cyclic FMD was the induction of atrophy and quiescence followed by vigorous regeneration and stem cell activation in several tissues. Similarly, metabolic parameters such as blood glucose, insulin, IGF-1, and IGF-1BP were reduced during the FMD phase and returned to control levels during a regular feeding regimen in both mice and human subjects (*27*, *28*).

There is substantial clinical interest in FMDs based on the rationale that their cyclic nature will increase compliance relative to traditional diets. In one randomized controlled crossover study, trimonthly FMD cycles of 5 days each reduced body mass index (BMI), fasting blood glucose, and blood pressure in obese, prediabetic subjects and subjects with hypertension, respectively (*28*). Cycles of FMD appear potentially beneficial in other clinical contexts, such as multiple sclerosis, autoimmune diseases, and cancer (*29*–*31*). Use of FMDs to improve outcomes in combination with chemotherapy is a particularly active area of research, with initial studies indicating that FMDs increase tumor sensitivity to chemotherapy in mouse models of breast cancer and melanoma while reducing collateral toxicity to healthy cells (*30*, *31*). However, a recent clinical trial failed to detect any difference in the specificity of chemotherapy in breast cancer patients undergoing FMD cycles (*32*), possibly because of low compliance (*33*).

Like IF, most FMD studies have been performed under conditions in which the experimental group consumed fewer calories than did the control group, and the extent to which isocaloric IF or FMDs have a substantial impact on life span or age-related pathology remains unclear. Some FMD studies report that FMD and control mice consume energetically equivalent amounts of food when normalized to body weight (*27*), but because the FMD mice weigh less than control mice, their total caloric intake is likely reduced. There is evidence that true isocaloric IF implemented as alternating feeding and fasting days (1:1 IF) is sufficient to induce ketogenesis during the fasting day and may improve metabolic homeostasis, stress resistance, and markers of inflammation compared with daily pair-fed mice (*34*). A 2-day feeding/1-day fasting isocaloric regimen (2:1 IF) also prevented weight and fat gain, maintained glucose and insulin homeostasis, and reduced adipocyte hypertrophy in mice fed an obesogenic diet, despite comparable energy intake. However, a recent 1:1 IF study over 3 weeks in lean, healthy people found that IF was less effective than isocaloric daily energy restriction and showed no benefits for measures of metabolic regulation or cardiovascular health (*35*). Because all of these studies were limited in scope and duration, future studies will need to assess whether isocaloric IF or FMDs have substantial long-term benefits on health and longevity in either rodents or people.

### *TRF*

TRF can be considered as a variant of IF in which subjects receive food every day but only for a specified time window. Isocaloric TRF studies in rodents suggest improvements in several metabolic parameters, including glucose and insulin homeostasis, energy expenditure, liver pathology, and resistance to different obesogenic diets (*36*–*38*). Intriguingly, isocaloric TRF seems to promote and maintain intrinsic circadian rhythms in mice (*36*, *37*), a phenotype also associated with CR (*39*). To the best of our knowledge, only one study has attempted to carefully examine the effect of isocaloric TRF on life span and age-related health outcomes in mice (*40*). In that report, which was limited to male mice, the TRF animals were trained to eat all of their food in a 12-hour window each day. The TRF mice were fed a diet intended to be isocaloric to the ad libitum group; however, the TRF animals still ended up eating less than did the ad libitum–fed animals. A 30% CR group in which the mice ate all of their food in a 3-hour window each day was also studied in parallel. The TRF group lived about 11% longer than did the ad libitum group, on average, whereas the 30% CR group showed a 28% increase in mean life span. Circulating levels of β-hydroxybutyrate were higher in the CR group but not in the TRF group.

Despite promising results of TRF in animal models, human studies are mixed. Some studies show only mild improvements (*41*), even when subjects naturally restricted themselves to 75 to 80% of their daily intake during feeding (*42*). Other studies indicate detrimental effects on glucose homeostasis (*43*). These studies designed their feeding window without regard for circadian variation. Larger, longer-term studies, carefully crafted around human circadian rhythms, are needed to determine whether TRF regimens can be beneficial to metabolic homeostasis and ultimately aging in humans.

### *PR and amino acid restriction*

The importance of protein as a dietary modulator of life span can be traced back to work in the 1920s that demonstrated that trout raised on a protein-deficient diet were both developmentally delayed and longer-lived (*44*). A few years later, dietary PR was found to delay development, sexual maturation, and signs of aging in rats (*45*). Since these early reports, numerous studies have described increased life span and reduced age-related pathology resulting from PR in rodents, which has been proposed to be mediated largely through reduced growth hormone, IGF, and mTOR signaling (*46*, *47*).

As with IF and FMDs, a particularly challenging aspect of the literature related to PR is disentangling the effects of reduced caloric intake from specific effects of protein itself as a dietary macronutrient. Speakman and colleagues performed a detailed meta-analysis comparing published effects of dietary interventions on life span in mice and rats and concluded that although there is evidence for life span extension from PR alone, the effects are substantially reduced compared with those reported associated with CR (*13*). This is consistent with work from Masoro and colleagues, who compared isocaloric PR to classic CR in rats and found that PR alone increased median life span by about 15% compared with about 50% increase from CR (*48*).

Nutritional geometry studies, carried out under ad libitum conditions in which ratios of different macronutrients are varied across numerous diets, provide an independent line of evidence that dietary protein may have an outsized impact on longevity in insects and mice, relative to other macronutrients (*49*). In mice, for example, a comprehensive study of longevity effects across 25 different diets found that those with lower protein-to-carbohydrate ratios yield the longest maximum life spans (*50*). The authors concluded that longevity and health are optimized in mice when protein is replaced with carbohydrate. However, the absolute life spans of the mice in this study were generally lower than those reported elsewhere for the same strain background (C57BL/6J), and the lowest-energy diets failed to yield the longest life spans (*50*). Further, out of all the diets studied, the one that resulted in the longest median life span (139 weeks) was quite high in protein (42% protein, 29% carbs, and 29% fat). Thus, the relationship between dietary protein and longevity, at least under ad libitum conditions, appears to be quite complex and influenced both by other macronutrients and additional variables that are not yet understood.

In addition to reducing the availability of total dietary protein, there are several reports of life span extension from restriction of specific essential amino acids, which must come from the diet and cannot be synthesized endogenously. The earliest of these may be from Segall, who published in 1977 that restriction of dietary tryptophan delayed growth, reduced cancer, and increased life span in rats (*51*). Orentreich and colleagues later found that restriction of dietary methionine (or more properly, sulfur amino acids that include both cysteine and methionine) similarly increased life span in rats (*52*). These seminal discoveries went largely unappreciated for many years but have gained attention as others reproduced and extended these findings to yeast, worms, fruit flies, and mice (*53*). More recently, restriction of the dietary BCAAs leucine, valine, and isoleucine has also been found to increase life span and delay age-related frailty in both fruit flies (*54*) and mice (*55*). BCAA restriction appears to increase life span through the inhibition of mTOR signaling. Other studies show that restricting methionine and cysteine is key to promoting a protective DR-mediated stress response that is blocked with mTOR activation. This sulfur amino acid restriction appears to be mediated through a mechanism that involves increased production of hydrogen sulfide gas through the transsulfuration pathway (*56*). Unlike most of the other antiaging diets that include CR, mice restricted for particular amino acids appear to actually eat more food than control-fed animals but fail to gain weight (*57*, *58*).

Along with inhibition of mTOR, the hormone fibroblast growth factor 21 (FGF21) has emerged as an important mediator of longevity benefits from PR and perhaps also amino acid restriction. FGF21 is actively secreted in response to reduced dietary protein in both mice and humans and is required for metabolic improvements in response to PR (*59*). FGF21 overexpression in mice fed ad libitum is sufficient to robustly increase life span (*60*). There is also evidence that FGF21 can be similarly induced through methionine restriction (*61*) and KDs (*62*). It has been proposed that FGF21 modulates life span in mice primarily by reducing growth hormone and IGF-1 signaling in liver (*60*).

### Do antiaging diets work in the "real world"?

Fad diets spawned from legitimate scientific research are nothing new, and recent years have seen a notable infiltration of antiaging diets into mainstream society. The renowned CR researcher Roy Walford attempted to popularize CR in the 1980s with his book *The 120 Year Diet: How to Double Your Vital Years* (*63*). Walford's "CR Society" never grew beyond a small group of followers, likely because of the severe abstinence required to maintain a CR lifestyle, and Walford himself died at the age of 79, well short of the 120 years promised in the book title. Recently, however, several less stringent variations on CR have achieved greater popularity, including IF, TRF, PR, and KDs. Some researchers studying these nutritional interventions follow in Walford's footsteps by practicing various forms of CR themselves and making dietary recommendations to the general public. Absent results from carefully designed clinical trials, this raises thorny questions around safety, efficacy, and scientific integrity.

When considering the evidence for antiaging diets in humans, we address two primary considerations. First, how strong is the case that these interventions actually slow or reverse biological aging in people? Second, are there potential side effects that may offset any benefits for healthy longevity? Before delving into these topics, however, we must differentiate between health benefits from modulation of biological aging versus effects that derive from being anti-obesogenic. CR, PR, and KDs are each used clinically to induce weight loss (*64*), and there is no question that weight loss in obese individuals can reduce disease risk. One rationale in favor of recommending these diets to the general public is the perception that they are all likely healthier, at least in the short term, than the typical Western diet. These potential benefits in overweight individuals should not, however, be conflated with effects on biological aging.

Unfortunately, it is not currently possible to know whether CR-like diets affect biological aging in people (Box 2). Unlike mice, controlled studies would need to be performed over many years to assess long-term benefits for life span and health span in humans. The recent development of various "aging clocks" that accurately predict chronological age, and may soon be useful for predicting biological age, offer the possibility that this question may be addressable in the relatively near future (*65*). For now, however, the data remain correlative. One line of evidence often cited to support antiaging effects of CR in natural human populations comes from studies of Okinawans, who inhabit a small Japanese island and smaller islands in the surrounding archipelago where the indigenous population historically consumed about 20% fewer calories than did the population of mainland Japan. Traditional Okinawan diets are very low protein (9% of total calories) and high carbohydrate (85% of total calories) (*66*). Historically, Okinawans enjoyed the longest life expectancy at birth and highest centenarian prevalence in the world, with remarkably low rates of age-associated diseases, such as cancer, heart and cardiovascular disease, and diabetes (*67*). Another line of evidence for health benefits from CR comes from the Comprehensive Assessment of Long-Term Effects of Reducing Intake of Energy (CALERIE) studies. These are a series of controlled clinical trials in normal and overweight adults subjected to a 25% reduction in caloric intake over periods ranging from a few months to 2 years. The results of these studies were generally consistent with improved clinical biomarkers of health such as decreased weight, enhanced insulin sensitivity and glucose tolerance, and improvements in major cardiometabolic risk factors (*68*). The CALERIE data are further complemented by uncontrolled studies of people self-practicing CR. Data from individuals self-practicing CR are also consistent with improved age-related health measures, including reduced weight and fat mass, lower blood pressure and other markers of heart disease, and improved glucose tolerance and insulin action (*69*, *70*).

Despite these suggestive data, there are concerns that laboratory nutrition studies may introduce artifacts that limit translation to humans. One example of this is that laboratory strains of rodents have been subjected to strong selection for rapid growth and early reproduction, which may lead to sensitization to the life span–extending effects of CR on

**Box 2. Common fictions about antiaging diets.**

Fiction: CR always "works." Although there are many reports of life span and health span extension from CR, there are also multiple published examples in which CR failed to extend life span. Among these are studies of wild-derived mice (*71*) and studies of genetically inbred mouse strains (*72*). Of the two long-term studies in rhesus monkeys, one reported a substantial increase in life span, whereas the other failed to detect any significant change (*98*, *99*). Although negative results can be challenging to interpret, the efficacy of CR even in laboratory animals appears to be highly dependent on sex, genetic background, the level of restriction used, and other variables yet to be identified (Fig. 1).

Fiction: CR extends life span only by preventing cancer. Although CR has been shown in many studies to have potent anticancer effects in rodents, it also delays age-related declines in immune, brain, heart, muscle, kidney, reproductive, and other tissues (*5*). CR also extends life span in nonmammalian species that do not get cancer, such as budding yeast, fruit flies, and nematode worms (*6*).

Fiction: Individual macronutrients are "good" or "bad" for aging. Dietary composition, total caloric intake, and feeding interval all have the potential to affect longevity and health span. It is possible to extend life span in mice by limiting total caloric intake, limiting primarily carbohydrates, or limiting primarily protein or even specific amino acids (Table 1). The mechanisms underlying these effects are complex and still poorly understood, even in highly controlled environments.

Fiction: Antiaging diets are known to slow aging in people. Despite their recent popularization, there is not yet strong evidence that any of the antiaging diets studied in laboratory animals have substantial long-term health benefits in nonobese humans.

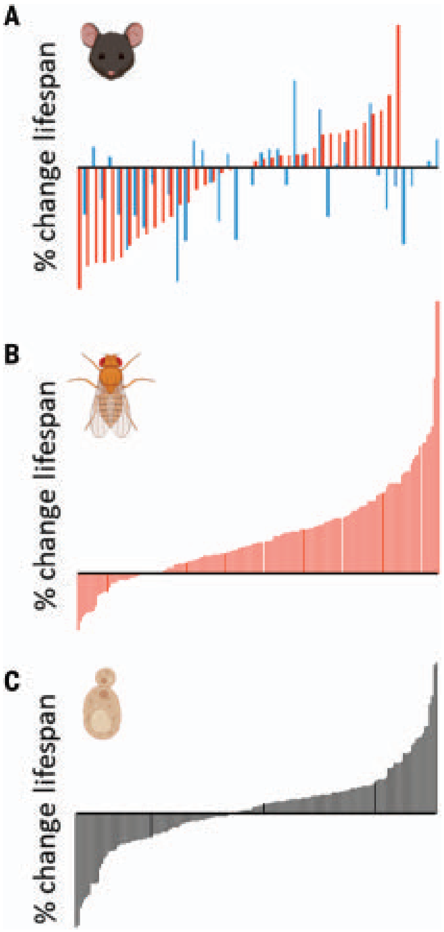


**Fig. 1. Percent life span change on restricted compared with unrestricted diet for genetically distinct strains of model organisms.** In every case, each bar along the *x* axis represents a different genetic background, and the percent change in life span of that strain in response to dietary restriction is shown on the *y* axis. (**A**) Change in mean life span for recombinant-inbred female (red) and male (blue) mice under ad libitum and 40% CR diets (*74*). (**B**) Change in median life span in genetically variable female flies from a natural population under a 10-fold PR (*77*). (**C**) Change in mean life span for single-gene-deletion mutant yeast under a 40-fold glucose restriction (*76*). Organism cartoons were created with BioRender (https://biorender.com).

growth-promoting signaling pathways such as growth hormone and mTOR (*71*). Another concern is that features of the laboratory environment may affect how animals respond to CR-like diets. Among these, mice and rats are typically housed in specific pathogen-free conditions and receive daily supervision and veterinary care as needed. Because of small body size and low temperatures at which they are housed, mice expend upward of half of the energy they consume just to maintain core body temperature (*72*). They experience consistent light-dark cycles and consume a refined, fixed composition chow for most of their adult life. None of these things are true in people, and it seems plausible that human environmental variation could have a large impact on health outcomes from dietary intervention. For example, severe CR can impair both immune function and wound healing (*73*), which could offset any potential life span–extending benefits under adverse environmental conditions in which the immune system is challenged (for example, a global viral pandemic) or in the absence of quality health care. Further complicating this interaction is the large variety in nutritional quality of human dietary composition, even among individuals who are consuming comparable total calories or macronutrient ratios.

The impact of genetic background may also limit translation from laboratory models to humans. Although not widely appreciated, genetic variation plays a large role in determining individual responses to CR in laboratory animals (Fig. 1). In one study, for example, a 40% reduction in caloric intake among recombinant inbred mouse lines significantly shortened life span in a greater number of genetic backgrounds than it extended (*74*). Sex-specific differences were also apparent, with CR having opposite effects on life span between sexes in some strains. Although this particular study was underpowered at the individual genotype level (*74*), several others have reported large variation in the response to CR in individual mouse and rat strains (*75*), and very similar genotype-dependent effects of DR by glucose restriction have been observed in budding yeast (*76*). The role of genetic variation in efficacy of other antiaging diets is largely unexplored, but one recent study found that about one-third of fruit fly strains experienced life span shortening in response to PR (*77*). It seems clear that a more detailed understanding of the mechanisms underlying variable response to antiaging diets in laboratory and genetically diverse animals will be important for predicting both efficacy and risk in people.

Differences in life span and age-associated nutritional requirements between humans and laboratory rodents present yet another layer of complexity when considering translation of antiaging diets. Although most animal studies examine the effects of lifelong nutritional intervention, very few people will maintain a CR (or PR or KD) lifestyle continuously over many decades of adulthood. Instead, repeated cycles of ad libitum and CR consumption is the norm. Detrimental effects of so-called "yo-yo dieting" are well documented and, taken to extremes, can result in a potentially fatal refeeding syndrome with severe consequences, such as hypotension, kidney injury, and heart failure (*78*). Even moderate dietary interventions are not without some level of risk over the long term. For example, one study found that low-protein diets are associated with reduced mortality in young people but higher mortality in people over the age of 65 years (*79*).

These correlative observations should provide a cautionary note. Although many people tend to assume that dietary interventions are safe, the biological effects of these antiaging diets are profound and generally less specific than pharmacological interventions. Like any drug, dietary interventions have a dose-response profile and at high enough "doses" will lead to substantial adverse effects and ultimately death. Among the potential side effects of CR-like diets are poor thermotolerance, loss of libido and sexual dysfunction, psychological problems, chronic fatigue, poor sleep, muscle weakness, susceptibility to infection, impaired wound healing, and social isolation (*80*). There is a very real likelihood that any given CR-like diet could enhance longevity in some people while shortening life span in others. The optimal nutritional strategy for longevity will certainly be different in different people, and there are few studies that quantify short-term or long-term side effects of antiaging diets in adults.

### mTOR inhibition: A common mechanism for antiaging diets?

Although many questions remain, research on antiaging diets has had a large impact on the field by providing insight into fundamental mechanisms of aging. The physiological consequences of dietary interventions are complex and multifactorial, even in relatively simple laboratory model organisms. Despite this, intriguing similarities across the spectrum of antiaging diets in diverse model systems have emerged that suggest at least partially overlapping mechanisms of action. As alluded to above, these mechanisms appear to converge on key nodes in a highly conserved aging-regulatory network (*7*, *8*). Although several components of this network have been implicated in mediating aspects of CR (*10*), a case can be made that mTOR is a particularly relevant and robust molecular transducer of diet-induced antiaging signals (*81*).

The mTOR protein is a kinase that mediates nutrient response signaling by acting in two distinct complexes, mTORC1 and mTORC2. Both complexes are affected by nutrient and growth cues, but mTORC1 has received the most scrutiny owing to observations that its inhibition is sufficient to mimic the effects of CR on life span in yeast, worms, flies, and mice (*82*). This has been replicated both genetically and pharmacologically (rapamycin) in each of these model systems. Like CR, mTORC1 inhibition appears to broadly delay or reverse age-related phenotypes across multiple tissues in mice, including the brain, heart, liver, kidney, immune system, skeletal muscle, auditory system, adipose, ovaries, and the oral cavity (*82*–*84*).

There is accumulating evidence that each of the antiaging diets discussed here can inhibit mTORC1 signaling through both direct and indirect mechanisms (Fig. 2). Specifically, mTORC1 is directly activated by leucine, an effect mediated in part by the sestrin family of proteins that inhibit mTORC1 only when not bound to leucine (*85*). Glucocorticoid signaling, which is induced by CR (*86*), also has an inhibitory effect on mTORC1 (*87*). mTORC1 is activated by the growth-promoting hormone IGF-1 through Akt and the tuberous sclerosis complex (TSC), and reduced IGF-1 signaling is one of the classic hallmarks of CR and CR-like diets (*5*). AMPK becomes activated in response to CR and regulates mTORC1 through phosphorylation of TSC (*81*). Several processes are proposed to mediate the life span–extending effects of CR downstream of mTORC1 inhibition. Covered in detail elsewhere, these include activation of autophagy, inhibition of mRNA translation, enhanced mitochondrial function, increased ketogenesis, improved stem cell function, and attenuation of senescence-associated inflammation (*81*, *88*, *89*).

We recognize that many questions remain regarding the relative importance of mTORC1 in mediating effects of CR and other antiaging interventions, and others have suggested that other factors such as sirtuins and AMPK are equally important (*90*). As described above, we favor the model that these factors all act together within a complex network that has yet to be fully characterized. Our focus here on mTORC1 reflects that it appears to be a particularly useful node within this network for modulating aging, as evidenced by the robust and reproducible data that support beneficial effects on life span and health span from pharmacological or genetic inhibition in yeast, worms, flies, and mice (*81*, *82*). We recognize that mTORC1 regulation is extremely complex, with tissue-dependent and circadian components that are still poorly understood. It is also clear that CR and mTOR inhibition have overlapping but distinct physiological effects so that treatment with rapamycin, for example, does not recapitulate all of the effects of CR, and vice versa (*91*). We are not aware of genetic backgrounds that fail to show life span extension in response to mTORC1 inhibition in mice, and unlike the case for CR, there have been no studies yet broadly assessing this question. High doses of rapamycin used to treat organ transplant patients are associated with multiple side effects, and although side effects are greatly reduced at lower doses in healthy people, whether mTORC1 inhibition is a useful therapeutic strategy to combat aging in humans remains unclear. Undoubtedly, much remains to be understood regarding the various interactions between dietary nutrients, longevity pathways, and healthy aging.

### Conclusion

Research on antiaging dietary interventions that increase life span and health span in laboratory models has greatly facilitated our mechanistic understanding of biological aging. Evolutionarily conserved signaling pathways

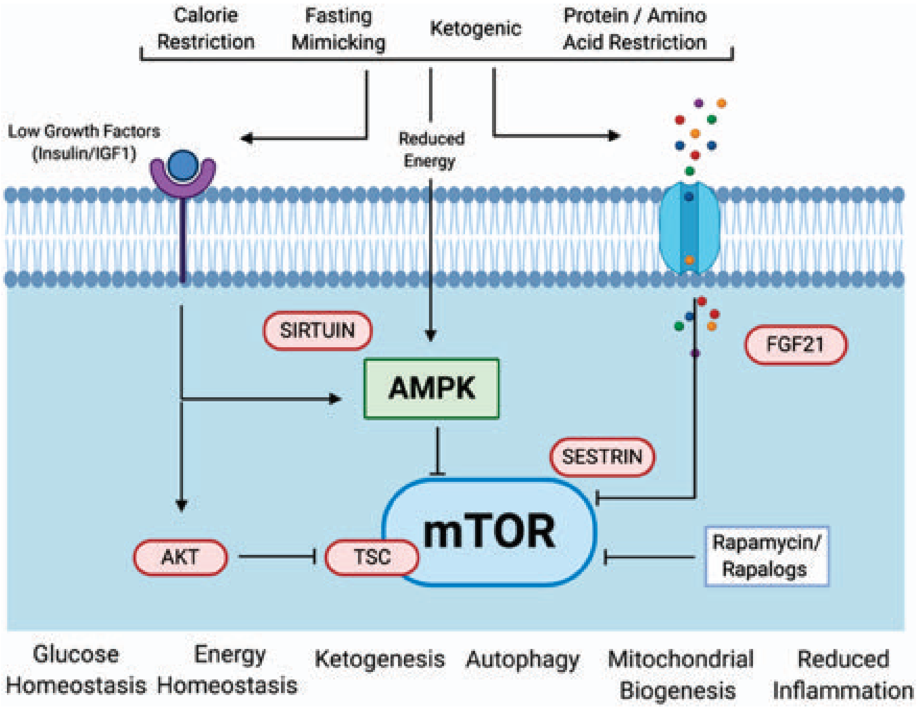


**Fig. 2. Diet modalities, molecular mechanisms, and downstream consequences of antiaging diets.** Dietary interventions that affect aging in mice limit one or more of the major dietary macromolecules and elicit cellular responses through a complex nutrient-sensing network. Key components of this network that have been implicated in effects on life span and health span in various laboratory model organisms include mTOR, FGF21, AMPK, insulin and IGF-1 receptors, AK strain transforming (AKT), sestrin, and sirtuins. The figure was created with BioRender (https://biorender.com).

appear to mediate many of the overlapping effects of antiaging diets, and their study has provided molecular targets for pharmacological interventions that may prove useful for increasing healthy longevity and reducing disease burden in humans. Although these diets have already achieved mainstream popularity in some cases, many questions remain about individual outcomes and relative risks associated with their long-term implementation. Future research should focus both on better understanding the cellular and molecular mediators of antiaging diets under highly controlled laboratory conditions as well as the impact of genetic and environmental variation on health outcomes associated with these diets.

## REFERENCES AND NOTES

1. C. M. McCay, M. F. Crowell, L. A. Maynard, The effect of retarded growth upon the length of life and upon ultimate size. *J. Nutr.* **10**, 63–79 (1935). doi: 10.1093/jn/10.1.63
2. T. B. Osborne, L. B. Mendel, E. L. Ferry, The effect of retardation of growth upon the breeding period and duration of life of rats. *Science* **45**, 294–295 (1917). doi: 10.1126/science.45.1160.294; pmid: 17760202
3. R. Weindruch, R. L. Walford, S. Fligiel, D. Guthrie, The retardation of aging in mice by dietary restriction: Longevity, cancer, immunity and lifetime energy intake. *J. Nutr.* **116**, 641–654 (1986). doi: 10.1093/jn/116.4.641; pmid: 3958810
4. J. R. Speakman, S. E. Mitchell, Caloric restriction. *Mol. Aspects Med.* **32**, 159–221 (2011). doi: 10.1016/j.mam.2011.07.001; pmid: 21840335
5. R. Weindruch, R. L. Walford, *The Retardation of Aging and Disease by Dietary Restriction* (C.C. Thomas, 1988), pp. xvii, 436 p.
6. P. Kapahi, M. Kaeberlein, M. Hansen, Dietary restriction and lifespan: Lessons from invertebrate models. *Ageing Res. Rev.* **39**, 3–14 (2017). doi: 10.1016/j.arr.2016.12.005; pmid: 28007498
7. T. Komatsu *et al*., Mechanisms of calorie restriction: A review of genes required for the life-extending and tumor-inhibiting effects of calorie restriction. *Nutrients* **11**, 3068 (2019). doi: 10.3390/nu11123068; pmid: 31888201
8. L. Fontana, L. Partridge, V. D. Longo, Extending healthy life span—From yeast to humans. *Science* **328**, 321–326 (2010). doi: 10.1126/science.1172539; pmid: 20395504
9. T. Chiba *et al*., Development of calorie restriction mimetics as therapeutics for obesity, diabetes, inflammatory and neurodegenerative diseases. *Curr. Genomics* **11**, 562–567 (2010). doi: 10.2174/138920210793360934; pmid: 21629433
10. F. Madeo, D. Carmona-Gutierrez, S. J. Hofer, G. Kroemer, Caloric restriction mimetics against age-associated disease: Targets, mechanisms, and therapeutic potential. *Cell Metab.* **29**, 592–610 (2019). doi: 10.1016/j.cmet.2019.01.018; pmid: 30840912
11. D. E. Harrison *et al*., Acarbose, 17-α-estradiol, and nordihydroguaiaretic acid extend mouse lifespan preferentially in males. *Aging Cell* **13**, 273–282 (2014). doi: 10.1111/acel.12170; pmid: 24245565
12. R. Strong *et al*., Longer lifespan in male mice treated with a weakly estrogenic agonist, an antioxidant, an α-glucosidase inhibitor or a Nrf2-inducer. *Aging Cell* **15**, 872–884 (2016). doi: 10.1111/acel.12496; pmid: 27312235
13. J. R. Speakman, S. E. Mitchell, M. Mazidi, Calories or protein? The effect of dietary restriction on lifespan in rodents is explained by calories alone. *Exp. Gerontol.* **86**, 28–38 (2016). doi: 10.1016/j.exger.2016.03.011; pmid: 27006163
14. R. M. Anderson, R. Weindruch, The caloric restriction paradigm: Implications for healthy human aging. *Am. J. Hum. Biol.* **24**, 101–106 (2012). doi: 10.1002/ajhb.22243; pmid: 22290875
15. D. K. Ingram, R. de Cabo, Calorie restriction in rodents: Caveats to consider. *Ageing Res. Rev.* **39**, 15–28 (2017). doi: 10.1016/j.arr.2017.05.008; pmid: 28610949
16. L. P. Sampaio, Ketogenic diet for epilepsy treatment. *Arq. Neuropsiquiatr.* **74**, 842–848 (2016). doi: 10.1590/0004-282X20160116; pmid: 27759811
17. K. D. Hall *et al*., Energy expenditure and body composition changes after an isocaloric ketogenic diet in overweight and obese men. *Am. J. Clin. Nutr.* **104**, 324–333 (2016). doi: 10.3945/ajcn.116.133561; pmid: 27385608
18. R. C. Atkins, *Dr. Atkins' Diet Revolution; The High Calorie Way to Stay Thin Forever* (D. McKay, 1972), pp. x, 310 p.
19. J. C. Newman *et al*., Ketogenic diet reduces midlife mortality and improves memory in aging mice. *Cell Metab.* **26**, 547–557.e8 (2017). doi: 10.1016/j.cmet.2017.08.004; pmid: 28877458
20. M. N. Roberts *et al*., A ketogenic diet extends longevity and healthspan in adult mice. *Cell Metab.* **26**, 539–546.e5 (2017). doi: 10.1016/j.cmet.2017.08.005; pmid: 28877457
21. J. C. Newman, E. Verdin, Ketone bodies as signaling metabolites. *Trends Endocrinol. Metab.* **25**, 42–52 (2014). doi: 10.1016/j.tem.2013.09.002; pmid: 24140022
22. C. Edwards *et al*., D-β-hydroxybutyrate extends lifespan in *C. elegans*. *Aging* **6**, 621–644 (2014). doi: 10.18632/aging.100683; pmid: 25127866
23. Y. Kashiwaya *et al*., A ketone ester diet exhibits anxiolytic and cognition-sparing properties, and lessens amyloid and tau pathologies in a mouse model of Alzheimer's disease. *Neurobiol. Aging* **34**, 1530–1539 (2013). doi: 10.1016/j.neurobiolaging.2012.11.023; pmid: 23276384
24. S. Srivastava *et al*., Mitochondrial biogenesis and increased uncoupling protein 1 in brown adipose tissue of mice fed a ketone ester diet. *FASEB J.* **26**, 2351–2362 (2012). doi: 10.1096/fj.11-200410; pmid: 22362892
25. M. P. Mattson, V. D. Longo, M. Harvie, Impact of intermittent fasting on health and disease processes. *Ageing Res. Rev.* **39**, 46–58 (2017). doi: 10.1016/j.arr.2016.10.005; pmid: 27810402
26. R. de Cabo, M. P. Mattson, Effects of intermittent fasting on health, aging, and disease. *N. Engl. J. Med.* **381**, 2541–2551 (2019). doi: 10.1056/NEJMra1905136; pmid: 31881139
27. S. Brandhorst *et al*., A periodic diet that mimics fasting promotes multi-system regeneration, enhanced cognitive performance, and healthspan. *Cell Metab.* **22**, 86–99 (2015). doi: 10.1016/j.cmet.2015.05.012; pmid: 26094889
28. M. Wei *et al*., Fasting-mimicking diet and markers/risk factors for aging, diabetes, cancer, and cardiovascular disease. *Sci. Transl. Med.* **9**, eaai8700 (2017). doi: 10.1126/scitranslmed.aai8700; pmid: 28202779
29. I. Y. Choi *et al*., A diet mimicking fasting promotes regeneration and reduces autoimmunity and multiple sclerosis symptoms. *Cell Rep.* **15**, 2136–2146 (2016). doi: 10.1016/j.celrep.2016.05.009; pmid: 27239035
30. I. Caffa *et al*., Fasting-mimicking diet and hormone therapy induce breast cancer regression. *Nature* **583**, 620–624 (2020). doi: 10.1038/s41586-020-2502-7; pmid: 32669709
31. S. Di Biase *et al*., Fasting-mimicking diet reduces HO-1 to promote T cell-mediated tumor cytotoxicity. *Cancer Cell* **30**, 136–146 (2016). doi: 10.1016/j.ccell.2016.06.005; pmid: 27411588
32. S. de Groot *et al*., Fasting mimicking diet as an adjunct to neoadjuvant chemotherapy for breast cancer in the multicentre randomized phase 2 DIRECT trial. *Nat. Commun.* **11**, 3083 (2020). doi: 10.1038/s41467-020-16138-3; pmid: 32576828
33. C. Vernieri, F. Ligorio, E. Zattarin, L. Rivoltini, F. de Braud, Fasting-mimicking diet plus chemotherapy in breast cancer treatment. *Nat. Commun.* **11**, 4274 (2020). doi: 10.1038/s41467-020-18194-1; pmid: 32848145
34. R. M. Anson *et al*., Intermittent fasting dissociates beneficial effects of dietary restriction on glucose metabolism and neuronal resistance to injury from calorie intake. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6216–6220 (2003). doi: 10.1073/pnas.1035720100; pmid: 12724520
35. I. Templeman *et al*., A randomized controlled trial to isolate the effects of fasting and energy restriction on weight loss and metabolic health in lean adults. *Sci. Transl. Med.* **13**, eabd8034 (2021). doi: 10.1126/scitranslmed.abd8034; pmid: 34135111
36. M. Hatori *et al*., Time-restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. *Cell Metab.* **15**, 848–860 (2012). doi: 10.1016/j.cmet.2012.04.019; pmid: 22608008
37. N. Martinez-Lopez *et al*., System-wide benefits of intermeal fasting by autophagy. *Cell Metab.* **26**, 856–871.e5 (2017). doi: 10.1016/j.cmet.2017.09.020; pmid: 29107505
38. A. Chaix, A. Zarrinpar, P. Miu, S. Panda, Time-restricted feeding is a preventative and therapeutic intervention against diverse nutritional challenges. *Cell Metab.* **20**, 991–1005 (2014). doi: 10.1016/j.cmet.2014.11.001; pmid: 25470547
39. O. Froy, R. Miskin, The interrelations among feeding, circadian rhythms and ageing. *Prog. Neurobiol.* **82**, 142–150 (2007). doi: 10.1016/j.pneurobio.2007.03.002; pmid: 17482337
40. S. J. Mitchell *et al*., Daily fasting improves health and survival in male mice independent of diet composition and calories. *Cell Metab.* **29**, 221–228.e3 (2019). doi: 10.1016/j.cmet.2018.08.011; pmid: 30197301
41. K. S. Stote *et al*., A controlled trial of reduced meal frequency without caloric restriction in healthy, normal-weight, middle-aged adults. *Am. J. Clin. Nutr.* **85**, 981–988 (2007). doi: 10.1093/ajcn/85.4.981; pmid: 17413096
42. S. Cienfuegos *et al*., Effects of 4- and 6-h time-restricted feeding on weight and cardiometabolic health: A randomized controlled trial in adults with obesity. *Cell Metab.* **32**, 366–378.e3 (2020). doi: 10.1016/j.cmet.2020.06.018; pmid: 32673591
43. O. Carlson *et al*., Impact of reduced meal frequency without caloric restriction on glucose regulation in healthy, normal-weight middle-aged men and women. *Metabolism* **56**, 1729–1734 (2007). doi: 10.1016/j.metabol.2007.07.018; pmid: 17998028
44. C. M. McCay, F. C. Bing, W. E. Dilley, Factor H in the nutrition of trout. *Science* **67**, 249–250 (1928). doi: 10.1126/science.67.1731.249; pmid: 17809873
45. J. Slonaker, Rollin, The effect of different per cents of protein in the diet VII. Life span and cause of death. *Am. J. Physiol.* **98**, 266–275 (1931). doi: 10.1152/ajplegacy.1931.98.2.266
46. R. K. Minor, J. S. Allard, C. M. Younts, T. M. Ward, R. de Cabo, Dietary interventions to extend life span and health span based on calorie restriction. *J. Gerontol. A Biol. Sci. Med. Sci.* **65**, 695–703 (2010). doi: 10.1093/gerona/glq042; pmid: 20371545
47. H. Mirzaei, R. Raynes, V. D. Longo, The conserved role of protein restriction in aging and disease. *Curr. Opin. Clin. Nutr. Metab. Care* **19**, 74–79 (2016). doi: 10.1097/MCO.0000000000000239; pmid: 26560522
48. B. P. Yu, E. J. Masoro, C. A. McMahan, Nutritional influences on aging of Fischer 344 rats: I. Physical, metabolic, and longevity characteristics. *J. Gerontol.* **40**, 657–670 (1985). doi: 10.1093/geronj/40.6.657; pmid: 4056321
49. S. J. Simpson *et al*., Dietary protein, aging and nutritional geometry. *Ageing Res. Rev.* **39**, 78–86 (2017). doi: 10.1016/j.arr.2017.03.001; pmid: 28274839
50. S. M. Solon-Biet *et al*., The ratio of macronutrients, not caloric intake, dictates cardiometabolic health, aging, and longevity in ad libitum-fed mice. *Cell Metab.* **19**, 418–430 (2014). doi: 10.1016/j.cmet.2014.02.009; pmid: 24606899
51. P. Segall, Long-term tryptophan restriction and aging in the rat. *Aktuelle Gerontol.* **7**, 535–538 (1977). pmid: 22255
52. N. Orentreich, J. R. Matias, A. DeFelice, J. A. Zimmerman, Low methionine ingestion by rats extends life span. *J. Nutr.* **123**, 269–274 (1993). pmid: 8429371
53. C. Hine, J. R. Mitchell, Calorie restriction and methionine restriction in control of endogenous hydrogen sulfide production by the transsulfuration pathway. *Exp. Gerontol.* **68**, 26–32 (2015). doi: 10.1016/j.exger.2014.12.010; pmid: 25523462
54. J. Lu *et al*., Sestrin is a key regulator of stem cell function and lifespan in response to dietary amino acids. *Nat. Aging* **1**, 60–72 (2020). doi: 10.1038/s43587-020-00001-7
55. N. E. Richardson *et al*., Lifelong restriction of dietary branched-chain amino acids has sex-specific benefits for frailty and lifespan in mice. *Nat. Aging* **1**, 73–86 (2021). doi: 10.1038/s43587-020-00006-2; pmid: 33796866
56. C. Hine *et al*., Endogenous hydrogen sulfide production is essential for dietary restriction benefits. *Cell* **160**, 132–144 (2015). doi: 10.1016/j.cell.2014.11.048; pmid: 25542313
57. B. E. Hasek *et al*., Dietary methionine restriction enhances metabolic flexibility and increases uncoupled respiration in both fed and fasted states. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **299**, R728–R739 (2010). doi: 10.1152/ajpregu.00837.2009; pmid: 20538896
58. L. Fontana *et al*., Decreased consumption of branched-chain amino acids improves metabolic health. *Cell Rep.* **16**, 520–530 (2016). doi: 10.1016/j.celrep.2016.05.092; pmid: 27346343
59. T. Laeger *et al*., FGF21 is an endocrine signal of protein restriction. *J. Clin. Invest.* **124**, 3913–3922 (2014). doi: 10.1172/JCI74915; pmid: 25133427
60. Y. Zhang *et al*., The starvation hormone, fibroblast growth factor-21, extends lifespan in mice. *eLife* **1**, e00065 (2012). doi: 10.7554/eLife.00065; pmid: 23066506
61. E. K. Lees *et al*., Methionine restriction restores a younger metabolic phenotype in adult mice with alterations in fibroblast

growth factor 21. *Aging Cell* **13**, 817–827 (2014). doi: 10.1111/acel.12238; pmid: 24935677
62. M. K. Badman *et al*., Hepatic fibroblast growth factor 21 is regulated by PPARα and is a key mediator of hepatic lipid metabolism in ketotic states. *Cell Metab.* **5**, 426–437 (2007). doi: 10.1016/j.cmet.2007.05.002; pmid: 17550778
63. R. L. Walford, The 120 Year Diet: How to Double Your Vital Years (Simon & Schuster, 1987).
64. C. Koliaki *et al*., Defining the optimal dietary approach for safe, effective and sustainable weight loss in overweight and obese adults. *Healthcare* **6**, 73 (2018). doi: 10.3390/healthcare6030073; pmid: 29958395
65. A. Zhavoronkov, P. Mamoshina, Deep aging clocks: The emergence of AI-based biomarkers of aging and longevity. *Trends Pharmacol. Sci.* **40**, 546–549 (2019). doi: 10.1016/j.tips.2019.05.004; pmid: 31279569
66. B. J. Willcox *et al*., Caloric restriction, the traditional Okinawan diet, and healthy aging: The diet of the world's longest-lived people and its potential impact on morbidity and life span. *Ann. N. Y. Acad. Sci.* **1114**, 434–455 (2007). doi: 10.1196/annals.1396.037; pmid: 17986602
67. B. J. Willcox, D. C. Willcox, M. Suzuki, Demographic, phenotypic, and genetic characteristics of centenarians in Okinawa and Japan: Part 1-centenarians in Okinawa. *Mech. Ageing Dev.* **165** (Pt B), 75–79 (2017). doi: 10.1016/j.mad.2016.11.001; pmid: 27845177
68. J. L. Dorling *et al*., Effects of caloric restriction on human physiological, psychological, and behavioral outcomes: Highlights from CALERIE phase 2. *Nutr. Rev.* **79**, 98–113 (2021). doi: 10.1093/nutrit/nuaa085; pmid: 32940695
69. L. Fontana, T. E. Meyer, S. Klein, J. O. Holloszy, Long-term calorie restriction is highly effective in reducing the risk for atherosclerosis in humans. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6659–6663 (2004). doi: 10.1073/pnas.0308291101; pmid: 15096581
70. L. Fontana, S. Klein, J. O. Holloszy, Effects of long-term calorie restriction and endurance exercise on glucose tolerance, insulin action, and adipokine production. *Age* **32**, 97–108 (2010). doi: 10.1007/s11357-009-9118-z; pmid: 19904628
71. J. M. Harper, C. W. Leathers, S. N. Austad, Does caloric restriction extend life in wild mice? *Aging Cell* **5**, 441–449 (2006). doi: 10.1111/j.1474-9726.2006.00236.x; pmid: 17054664
72. V. Škop *et al*., Mouse thermoregulation: Introducing the concept of the thermoneutral point. *Cell Rep.* **31**, 107501 (2020). doi: 10.1016/j.celrep.2020.03.065; pmid: 32294435
73. J. K. Stechmiller, Understanding the role of nutrition and wound healing. *Nutr. Clin. Pract.* **25**, 61–68 (2010). doi: 10.1177/0884533609358997; pmid: 20130158
74. C. Y. Liao, B. A. Rikke, T. E. Johnson, V. Diaz, J. F. Nelson, Genetic variation in the murine lifespan response to dietary restriction: From life extension to life shortening. *Aging Cell* **9**, 92–95 (2010). doi: 10.1111/j.1474-9726.2009.00533.x; pmid: 19878144
75. W. R. Swindell, Dietary restriction in rats and mice: A meta-analysis and review of the evidence for genotype-dependent effects on lifespan. *Ageing Res. Rev.* **11**, 254–270 (2012). doi: 10.1016/j.arr.2011.12.006; pmid: 22210149
76. J. Schleit *et al*., Molecular mechanisms underlying genotype-dependent responses to dietary restriction. *Aging Cell* **12**, 1050–1061 (2013). doi: 10.1111/acel.12130; pmid: 23837470
77. K. Jin *et al*., Genetic and metabolomic architecture of variation in diet restriction-mediated lifespan extension in *Drosophila*. *PLOS Genet.* **16**, e1008835 (2020). doi: 10.1371/journal.pgen.1008835; pmid: 32644988
78. M. A. Crook, Refeeding syndrome: Problems with definition and management. *Nutrition* **30**, 1448–1455 (2014). doi: 10.1016/j.nut.2014.03.026; pmid: 25280426
79. M. E. Levine *et al*., Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population. *Cell Metab.* **19**, 407–417 (2014). doi: 10.1016/j.cmet.2014.02.006; pmid: 24606898
80. J. Most, V. Tosti, L. M. Redman, L. Fontana, Calorie restriction in humans: An update. *Ageing Res. Rev.* **39**, 36–45 (2017). doi: 10.1016/j.arr.2016.08.005; pmid: 27544442
81. B. K. Kennedy, D. W. Lamming, The mechanistic target of rapamycin: The grand conducTOR of metabolism and aging. *Cell Metab.* **23**, 990–1003 (2016). doi: 10.1016/j.cmet.2016.05.009; pmid: 27304501
82. S. C. Johnson, P. S. Rabinovitch, M. Kaeberlein, mTOR is a key modulator of ageing and age-related disease. *Nature* **493**, 338–345 (2013). doi: 10.1038/nature11861; pmid: 23325216
83. R. A. Altschuler *et al*., Rapamycin added to diet in late mid-life delays age-related hearing loss in UMHET4 mice. *Front. Cell. Neurosci.* **15**, 658972 (2021). doi: 10.3389/fncel.2021.658972; pmid: 33897373
84. J. Y. An *et al*., Rapamycin rejuvenates oral health in aging mice. *eLife* **9**, e54318 (2020). doi: 10.7554/eLife.54318; pmid: 32342860
85. R. L. Wolfson *et al*., Sestrin2 is a leucine sensor for the mTORC1 pathway. *Science* **351**, 43–48 (2016). doi: 10.1126/science.aab2674; pmid: 26449471
86. N. V. Patel, C. E. Finch, The glucocorticoid paradox of caloric restriction in slowing brain aging. *Neurobiol. Aging* **23**, 707–717 (2002). doi: 10.1016/S0197-4580(02)00017-9; pmid: 12392776
87. O. Schakman, S. Kalista, C. Barbé, A. Loumaye, J. P. Thissen, Glucocorticoid-induced skeletal muscle atrophy. *Int. J. Biochem. Cell Biol.* **45**, 2163–2172 (2013). doi: 10.1016/j.biocel.2013.05.036; pmid: 23806868
88. T. Weichhart, mTOR as regulator of lifespan, aging, and cellular senescence: A mini-review. *Gerontology* **64**, 127–134 (2018). doi: 10.1159/000484629; pmid: 29190625
89. S. C. Johnson, M. Sangesland, M. Kaeberlein, P. S. Rabinovitch, Modulating mTOR in aging and health. *Interdiscip. Top. Gerontol.* **40**, 107–127 (2015). doi: 10.1159/000364974; pmid: 25341517
90. J. A. Baur *et al*., Dietary restriction: Standing up for sirtuins. *Science* **329**, 1012–1013, author reply 1013–1014 (2010). doi: 10.1126/science.329.5995.1012; pmid: 20798296
91. W. C. Fok *et al*., Short-term treatment with rapamycin and dietary restriction have overlapping and distinctive effects in young mice. *J. Gerontol. A Biol. Sci. Med. Sci.* **68**, 108–116 (2013). doi: 10.1093/gerona/gls127; pmid: 22570137
92. K. Xie *et al*., Every-other-day feeding extends lifespan but fails to delay many symptoms of aging in mice. *Nat. Commun.* **8**, 155 (2017). doi: 10.1038/s41467-017-00178-3; pmid: 28761067
93. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The hallmarks of aging. *Cell* **153**, 1194–1217 (2013). doi: 10.1016/j.cell.2013.05.039; pmid: 23746838
94. B. K. Kennedy *et al*., Geroscience: Linking aging to chronic disease. *Cell* **159**, 709–713 (2014). doi: 10.1016/j.cell.2014.10.039; pmid: 25417146
95. M. Kaeberlein, How healthy is the healthspan concept? *Geroscience* **40**, 361–364 (2018). doi: 10.1007/s11357-018-0036-9; pmid: 30084059
96. R. Selvarani, S. Mohammed, A. Richardson, Effect of rapamycin on aging and age-related diseases-past and future. *Geroscience* **43**, 1135–1158 (2021). doi: 10.1007/s11357-020-00274-1; pmid: 33037985
97. M. B. Lee, M. Kaeberlein, Translational Geroscience: From invertebrate models to companion animal and human interventions. *Transl. Med. Aging* **2**, 15–29 (2018). doi: 10.1016/j.tma.2018.08.002; pmid: 32368707
98. R. J. Colman *et al*., Caloric restriction delays disease onset and mortality in rhesus monkeys. *Science* **325**, 201–204 (2009). doi: 10.1126/science.1173635; pmid: 19590001
99. R. J. Colman *et al*., Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. *Nat. Commun.* **5**, 3557 (2014). doi: 10.1038/ncomms4557; pmid: 24691430

## ACKNOWLEDGMENTS

The authors thank D. Promislow, T. Kaeberlein, and C. Morrison for careful review of the manuscript. All authors contributed to drafting the manuscript. The authors dedicate this work to the memory of James R. Mitchell, whose recent untimely passing could not overshadow his important contributions to the field of aging research. **Funding:** M.K. and A.B. are supported by the University of Washington Nathan Shock Center of Excellence in the Basic Biology of Aging (P30 AG013280). M.B.L. was supported by the National Institutes of Health (NIH) Alzheimer's Disease Training Program (T32 AG052354). C.M.H. was supported by the National Institutes of Health (K99AG070273) and the Pennington Biomedical NORC-Nutrition and Metabolic Health Through the Lifespan (P30 DK072476). **Competing interests:** We declare no competing interests.

10.1126/science.abe7365

RESEARCH ARTICLE SUMMARY

DEVELOPMENTAL BIOLOGY

# The X chromosome dosage compensation program during the development of cynomolgus monkeys

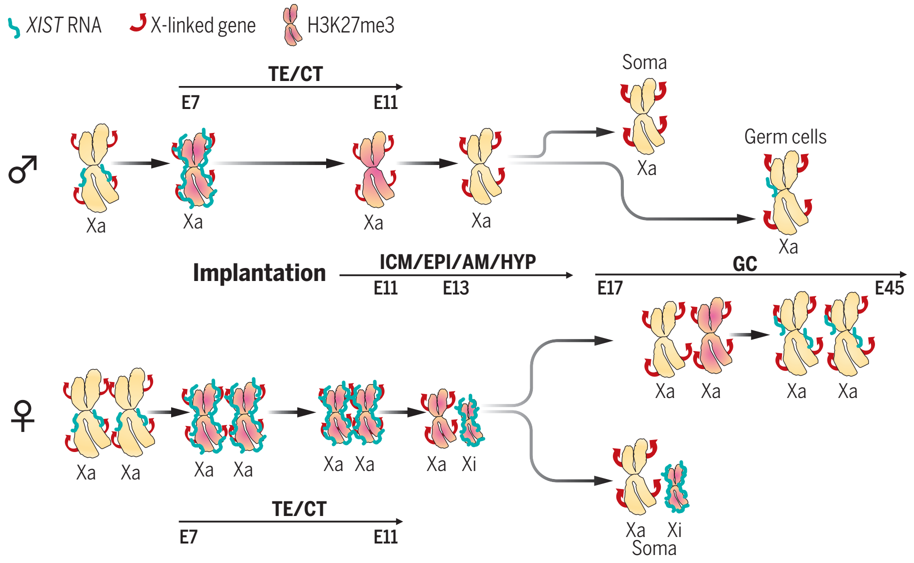
**Ikuhiro Okamoto*, Tomonori Nakamura, Kotaro Sasaki, Yukihiro Yabuta, Chizuru Iwatani, Hideaki Tsuchiya, Shin-ichiro Nakamura, Masatsugu Ema, Takuya Yamamoto, Mitinori Saitou***

**INTRODUCTION:** Mammals display dimorphic sex chromosomes, the X and Y chromosomes. Females have two Xs, whereas males have one X and one Y. The X encodes more than 800 genes; by contrast, the Y encodes the critical male determinant *SRY* but is generally gene-poor. Mechanisms have evolved to ensure a balanced gene dosage between the X and autosomes and between the sexes: X chromosome up-regulation (XCU) and inactivation (XCI). XCI has been studied intensively in mice, creating a paradigm for the epigenetic regulation of chromosomal activities. The long noncoding RNA *Xist* (X-inactive specific transcript) plays a central role in activating a cascade of repressive events on the X, including histone H3 lysine 27 trimethylation (H3K27me3) and histone H2A lysine 119 monoubiquitination (H2AK119u1), leading to XCI early during development. However, recent studies have uncovered critical divergences in the XCI mechanism among mammals. In humans, *XIST* is not sufficient for XCI, and the mechanism of XCI remains unclear.

**RATIONALE:** Cynomolgus monkeys (*Macaca fascicularis*) are primates that are closely related to humans. They could serve as an appropriate model for human development and provide key information regarding the evolution of the dosage compensation mechanisms. We performed a comprehensive characterization of X chromosome activities during early development and germ cell development in cynomolgus monkeys.

**RESULTS:** Female embryos expressed *XIST* from both paternal and maternal Xs. *XIST* coated both Xs in cis and induced the repressive modification H3K27me3 and H2AK119u1. This occurred in both the trophectoderm (TE) and inner cell mass (ICM) and was accompanied by compaction of the X chromosomes. However, until the blastocyst stage at embryonic day 9 (E9), a majority of such Xs remained active and did not initiate chromosome-wide XCI. On the Xs destined toward the active state, *XIST* was repressed and erasure of repressive modifications ensued, whereas on the Xs destined toward the inactive state, *XIST* up-regulation and enrichment of repressive modifications continued, condensing the whole chromosome for XCI. These events occurred in the embryonic and extraembryonic lineages, with earlier XCI establishment in the TE-derived cytotrophoblast around peri-implantation (about E11) and protracted XCI establishment in the ICM-derived amnion, epiblast, and hypoblast after implantation (about E15). The male single Xs carried out the same events: *XIST* coating, H3K27me3 and H2AK119u1 enrichment, and "compaction," followed by *XIST* repression and repressive-modification erasure around peri-implantation. Single-cell transcriptome analyses revealed that males achieve XCU progressively, whereas females show XCU coincidentally with XCI, both establishing the X:autosome dosage compensation by 1 week after implantation. On the other hand, female germ cells expressed X-linked genes biallelically without repression or with reactivation after their specification, repressed *XIST* on both Xs (E13–E17), and then erased H3K27me3 (E22–E28). Thereafter, they gained full biallelic X-linked gene expression, with *XIST* reexpression (E42).

**CONCLUSION:** We have elucidated a comprehensive scenario for the X chromosome dosage compensation during cynomolgus monkey development. The finding that the Xs with *XIST* coating, H3K27me3 and H2AK119u1 enrichment, and a compacted structure are an active intermediate that persists after implantation in the epiblast indicates the existence of an as yet unidentified mechanism(s) essential for triggering chromosome-wide XCI in primates. Our analysis involving the genes relevant to dosage compensation favors the notion that X chromosome dampening may not be a major mechanism for dosage compensation. Our study clarifies the ontogeny and a distinct mechanism of the X chromosome dosage compensation in cynomolgus monkeys, which show critical divergences from those in mice and bear direct relevance not only to human X chromosome biology, but also to human and primate female stem cell biology, in which an appropriate regulation of the Xs is a critical requirement. ■



**Dynamics of X chromosome activities during monkey development.** *XIST* (green lines) and X-linked gene (red arrows) expression and H3K27me3/H2AK119u1 enrichment (pink shade) on the X chromosomes, and X chromosome inactivation/reactivation during male (top) and female (bottom) cynomolgus monkey development. Xa, active X; Xi, inactive X; TE, trophectoderm; CT, cytotrophoblast; ICM, inner cell mass; EPI, epiblast; AM, amnion; HYP, hypoblast; GC, germ cells; E, embryonic day.

---

The list of author affiliations is available in the full article online.
*Corresponding author. Email: iokamoto@anat2.med.kyoto-u.ac.jp (I.O.); saitou@anat2.med.kyoto-u.ac.jp (M.S.)

Cite this article as I. Okamoto *et al.*, *Science* **374**, eabd8887 (2021). DOI: 10.1126/science.abd8887

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd8887

RESEARCH ARTICLE

DEVELOPMENTAL BIOLOGY

# The X chromosome dosage compensation program during the development of cynomolgus monkeys

Ikuhiro Okamoto[1,2]*, Tomonori Nakamura[1,2,3], Kotaro Sasaki[2]†, Yukihiro Yabuta[1,2],
Chizuru Iwatani[4], Hideaki Tsuchiya[4], Shin-ichiro Nakamura[4]‡, Masatsugu Ema[1,4],
Takuya Yamamoto[1,5,6,7], Mitinori Saitou[1,2,5]*

X chromosome dosage compensation ensures balanced gene dosage between the X chromosome and autosomes and between the sexes, involving divergent mechanisms among mammals. We elucidated a distinct mechanism for X chromosome inactivation (XCI) in cynomolgus monkeys, a model for human development. The trophectoderm and cytotrophoblast acquire XCI around implantation through an active intermediate bearing repressive modifications and compacted structure, whereas the amnion, epiblast, and hypoblast maintain such an intermediate protractedly, attaining XCI by a week after implantation. Males achieve X chromosome up-regulation (XCU) progressively, whereas females show XCU coincidentally with XCI, both establishing the X:autosome dosage compensation by 1 week after implantation. Conversely, primordial germ cells undergo X chromosome reactivation by reversing the XCI pathway early during their development. Our findings establish a foundation for clarifying the dosage compensation mechanisms in primates, including humans.

The therian mammals bear sexually dimorphic chromosomes: the X and Y chromosomes, which evolved from an ancestral autosomal pair (*1*). Females bear two X chromosomes, which encode >800 genes with critical physiological functions, whereas males have one X and one Y, the latter of which is gene-poor, encoding ~70 genes, including the male determinant *SRY* (*2*, *3*). X chromosome dosage mechanisms have evolved to ensure balanced gene dosage between the X and autosomes and between the sexes (*1*, *4*): the former compensation entails an up-regulation of X-linked gene products (X chromosome up-regulation: XCU) in both sexes (*1*, *5*), and the latter involves random silencing of one X (X chromosome inactivation: XCI) in females (*4*, *6*). A failure in XCI results in embryonic lethality (*7*), and both XCI and XCU contribute to disease states involving X-linked gene mutations (*8*).

XCI has been studied extensively in mice (*6*). During female mouse development, the long noncoding RNA (lncRNA) *Xist* (X-inactive specific transcript) (*9*–*11*) begins its expression from the paternal X (Xp) upon zygotic genome activation at the two-cell stage, coats the Xp in cis, and induces repressive modifications, including histone H3 $\text{Lys}^{27}$ trimethylation (H3K27me3) and histone H2A $\text{Lys}^{119}$ monoubiquitination (H2AK119u1)—mediated by polycomb repressive complex (PRC) 2 and 1, respectively—to silence the Xp from the four- to the eight-cell stage (*12*–*15*). The Xp remains silenced in the trophectoderm (TE) (*16*), whereas it becomes reactivated with *Xist* down-regulation and erasure of repressive modifications in the inner cell mass (ICM), and random XCI ensues in the epiblast after implantation (*13*, *14*). In males, *Xist* on the maternal X (Xm) is repressed by histone H3 $\text{Lys}^{9}$ trimethylation (H3K9me3) and H3K27me3 (*17*, *18*), and the Xm is active throughout development. Thus, in mice, *Xist* induces a hierarchy of repressive events for XCI. The in vitro system using female mouse embryonic stem cells (mESCs) has been instrumental in clarifying the underlying molecular mechanisms (*6*).

The XCI mechanisms differ among mammals (*19*). During human development, *XIST* is not imprinted, begins its expression from both Xp and Xm upon zygotic genome activation (eight-cell stage), and coats both Xs in cis, but H3K27me3 does not accumulate and chromosome-wide XCI does not begin by the blastocyst stage (*20*). Therefore, it appears that in humans, *XIST* is not sufficient to induce XCI. There is conflicting evidence as to whether *XIST* dampens the X chromosome activity (*21*–*24*). Recently, in vitro human blastocyst cultures have suggested that XCI begins soon after implantation (*25*, *26*), but the details of the XCI mechanism and its progression during human development remain unclear. Whether the in vitro cultures recapitulate the in vivo development also requires clarification. Moreover, the XCI properties in female human ESCs (hESCs) and induced pluripotent stem cells are markedly different from those in female mESCs, and such differences influence the phenotypes of their differentiated progenies (*19*, *27*, *28*). Thus, it is important to understand the regulation of X chromosome activities during a critical developmental window in humans and closely related nonhuman primates (*29*). Here, we performed a comprehensive characterization of X chromosome activities during development in cynomolgus monkeys (cy; *Macaca fascicularis*).

## Active Xs with *XIST* and repressive modifications in pre-implantation embryos

We examined the expression of *XIST* and X-linked genes (*POLA1*, *HUWE1*, *ATRX*) in pre-implantation embryos by RNA fluorescence in situ hybridization (RNA-FISH). *ATRX* locates close to *XIST* and is repressed rapidly upon XCI, whereas *POLA1* and *HUWE1* locate more distantly from *XIST* and are repressed relatively slowly upon XCI (*15*, *20*). We confirmed the probes using female cy embryonic fibroblasts (CEFs) (fig. S1). At the eight-cell stage [embryonic day 3 (E3)], both *XIST* and *POLA1* were detected as two and single punctate signals in all female and male blastomeres, respectively (fig. S2), indicating that they initiate expression from both Xs upon zygotic genome activation (*30*). In cy embryos, the epiblast (EPI) originates from ICM at E6, and the TE and hypoblast are specified at E6 and E7, respectively (*31*). In females, at E7 the majority of the TE showed two diffuse *XIST* domains with adjacent X-linked gene signals ($\mathrm{Xa}^{XIST+}/\mathrm{Xa}^{XIST+}$) (Fig. 1, A and B, and fig. S3A). At E8, ~70% of the TE retained two $\mathrm{Xa}^{XIST+}$, whereas ~20% showed one *XIST* and two X-linked gene signals ($\mathrm{Xa}^{XIST+}/\mathrm{Xa}^{XIST-}$) and fewer than ~10% exhibited one *XIST* and one X-linked gene signal ($\mathrm{Xi}^{XIST+}/\mathrm{Xa}^{XIST-}$) (Fig. 1, A and B). At E9, the ratio of $\mathrm{Xi}^{XIST+}/\mathrm{Xa}^{XIST-}$ TE increased slightly (Fig. 1, A and B, and fig. S3A). The $XIST^{+}$ domains showed minor enrichment of 4′,6-diamidino-2-phenylindole (DAPI) signals from E7 to E9 (fig. S4).

In the ICM isolated by immunosurgery at E8 (fig. S5), most cells were $\mathrm{Xa}^{XIST+}/\mathrm{Xa}^{XIST+}$, with a minor population of $\mathrm{Xa}^{XIST+}/\mathrm{Xa}^{XIST-}$ or $\mathrm{Xi}^{XIST+}/\mathrm{Xa}^{XIST-}$ cells (Fig. 1, C and D). We did not find $\mathrm{Xa}^{XIST-}/\mathrm{Xa}^{XIST-}$ cells either in the TE or ICM (Fig. 1, B and D). In males, ~60% of TE

---

[1] Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. [2] Department of Anatomy and Cell Biology, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. [3] Hakubi Center for Advanced Research, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. [4] Research Center for Animal Life Science, Shiga University of Medical Science, Seta-Tsukinowa-cho, Otsu, Shiga 520-2192, Japan. [5] Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan. [6] AMED-CREST, AMED, 1-7-1 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan. [7] Medical-Risk Avoidance Based on iPS Cells Team, RIKEN Center for Advanced Intelligence Project (AIP), Kyoto 606-8507, Japan.

***Corresponding author. Email: iokamoto@anat2.med.kyoto-u.ac.jp (I.O.); saitou@anat2.med.kyoto-u.ac.jp (M.S.)**

†Present address: Department of Biomedical Sciences, University of Pennsylvania School of Veterinary Medicine, Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA.

‡Present address: Laboratory of Laboratory Animal Science, School of Veterinary Medicine, Azabu University, 1-17-71 Fuchinobe, Chuo, Sagamihara, Kanagawa, Japan.

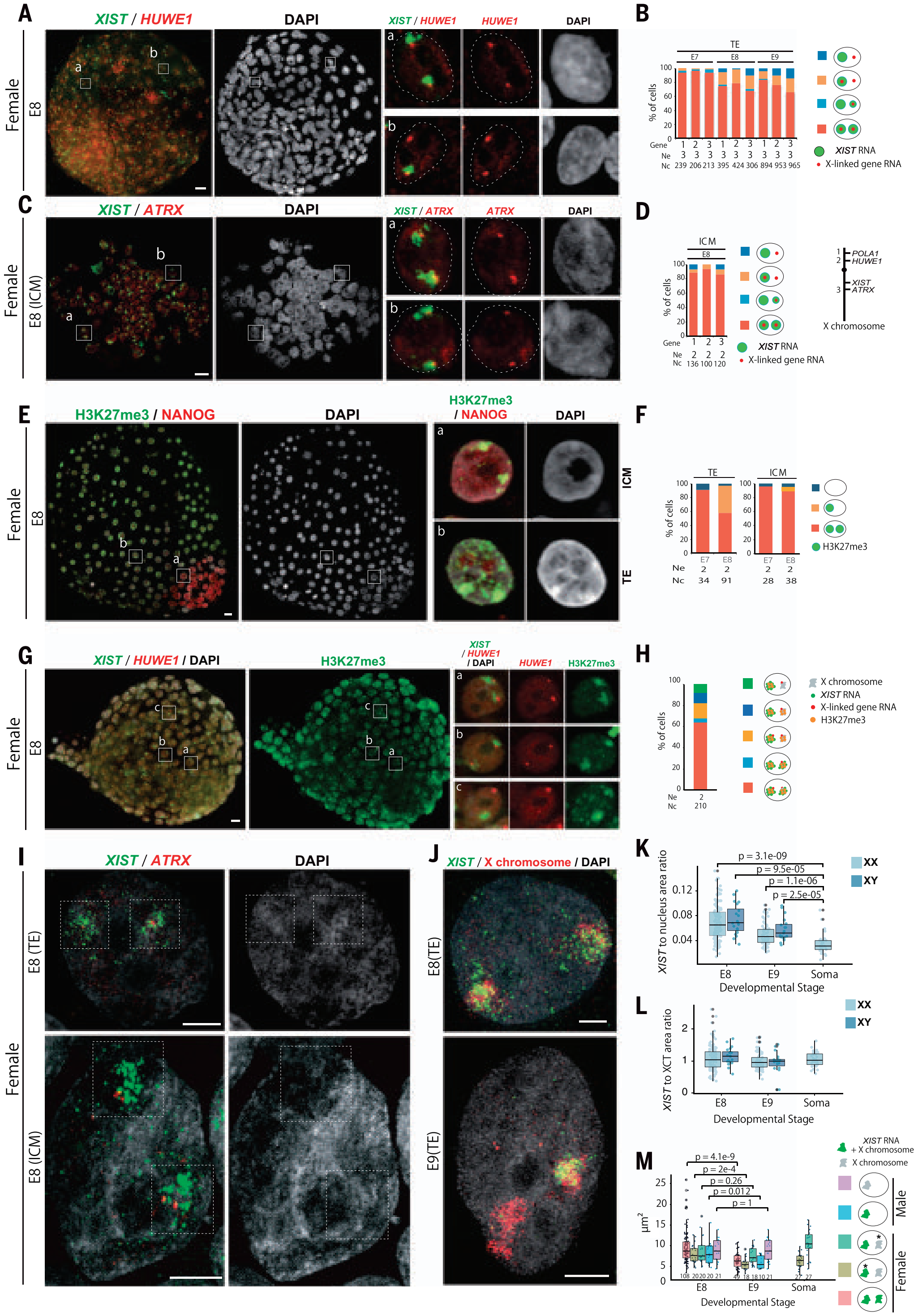

A
XIST / HUWE1
DAPI
XIST / HUWE1
HUWE1
DAPI
Female
E8
a
b
B
TE
E7
E8
E9
% of cells
Gene
Ne
Nc
1 2 3 1 2 3 1 2 3
3 3 3 3 3 3 3 3 3
239 206 213 395 424 306 894 953 965
XIST RNA
X-linked gene RNA
C
XIST / ATRX
DAPI
XIST / ATRX
ATRX
DAPI
Female
E8 (ICM)
D
ICM
E8
% of cells
Gene
Ne
Nc
1 2 3
2 2 2
136 100 120
XIST RNA
X-linked gene RNA
POLA1
HUWE1
XIST
ATRX
X chromosome
E
H3K27me3 / NANOG
DAPI
H3K27me3 / NANOG
DAPI
Female
E8
ICM
TE
F
TE
ICM
% of cells
Ne
Nc
E7 E8
2 2
34 91
E7 E8
2 2
28 38
H3K27me3
G
XIST / HUWE1 / DAPI
H3K27me3
XIST / HUWE1 / DAPI
HUWE1
H3K27me3
Female
E8
a
b
c
H
% of cells
Ne 2
Nc 210
X chromosome
XIST RNA
X-linked gene RNA
H3K27me3
I
XIST / ATRX
DAPI
Female
E8 (TE)
E8 (ICM)
J
XIST / X chromosome / DAPI
E8(TE)
E9(TE)
K
XIST to nucleus area ratio
p = 3.1e-09
p = 9.5e-05
p = 1.1e-06
p = 2.5e-05
XX
XY
E8
E9
Soma
Developmental Stage
L
XIST to XCT area ratio
XX
XY
E8
E9
Soma
Developmental Stage
M
µm²
p = 4.1e-9
p = 2e-4
p = 0.26
p = 0.012
p = 1
E8
E9
Soma
Developmental Stage
XIST RNA + X chromosome
X chromosome
Male
Female

**Fig. 1. X chromosome properties in female cy pre-implantation embryos.** (**A** and **C**) A female E8 embryo (A) or ICM (C) analyzed by RNA-FISH for *XIST* (green) and *HUWE1* [red in (A)] or *ATRX* [red in (C)]. DAPI staining is shown in gray, here and below. Two representative nuclei (boxed) are magnified. (**B** and **D**) Percentages of TE (E7–E9; females) (B) or ICM cells (E8; females) (D) showing distinct *XIST*/X-linked gene expression as indicated by schemes and color codes. In all relevant figure panels, approximate X-linked gene positions on the X are shown with their numbers assigned. Nc, number of cells analyzed; Ne, number of embryos analyzed. (**E**) A female E8 embryo with immunofluorescence for H3K27me3 (green) and NANOG (red). Two representative nuclei (boxed) (a, ICM; b, TE) are magnified. (**F**) Percentages of TE and ICM cells (E7–E8; females) showing distinct H3K27me3 enrichments as indicated by schemes and color codes. (**G**) A female E8 embryo analyzed by RNA-FISH for *XIST* (green) and *HUWE1* (red) followed by immunofluorescence with H3K27me3 (green). Three representative TE nuclei (boxed) are magnified. (**H**) Percentages of TE (E8) showing distinct *XIST*/*HUWE1* expression and H3K27me3 enrichments as indicated by schemes and color codes. (**I**) 3D-SIM images of *XIST* RNA (green) and *ATRX* (red) in a TE (top) or ICM cell (bottom) of female E8 embryos. Two X chromosomes (dotted boxes) are magnified. (**J**) 3D-SIM merged images of *XIST* RNA (green), X chromosome (red) and DAPI (gray) in the TE cells of female embryos (E8/9). (**K** to **M**) The *XIST* dispersion area (M), relative to the nuclear size (K) or the X chromosome territory (L), in female and male TE (E8–E9) and CEFs measured using 3D-SIM. The color code and the scheme for the X chromosome states are as indicated. Scale bars, 10 µm [(A), (C), (E), and (G)], 2 µm [(I) and (J)].

at E7 bore a single $Xa^{XIST+}$, but after E8, 80% or more showed a single $Xa^{XIST-}$, whereas ~80% of ICM cells at E8 had a single $Xa^{XIST+}$ (fig. S3, B to D). Thus, cy embryos express *XIST* from both Xp and Xm, and *XIST* coats them in cis, but chromosome-wide XCI does not begin by the blastocyst stage in a majority of TE and ICM cells (see below for global X-linked gene expression analyses).

We next examined whether the Xs might accumulate repressive epigenetic modifications. Immunofluorescence analyses in females revealed that, at E7, most NANOG$^-$ TE and NANOG$^+$ ICM cells showed two H3K27me3-enriched domains, which colocalized with relatively DAPI-dense foci (Fig. 1, E and F, and fig. S6A). At E8, about half of NANOG$^-$ TE cells lost one H3K27me3-enriched domain, whereas ~90% of NANOG$^+$ ICM cells retained two H3K27me3 domains (Fig. 1, E and F). In males, ~70% of TE cells at E7 bore a single H3K27me3 domain, but at E8, ~90% lost it, whereas ~30% of ICM cells retained a single H3K27me3 domain at both E7 and E8 (fig. S6, B and C). RNA-FISH followed by immunofluorescence demonstrated the colocalization of *XIST* and *HUWE1* with H3K27me3 (Fig. 1, G and H, and fig. S6D). In female TE at E8, a majority (>60%) of cells were Xa$^{XIST+}$/Xa$^{XIST+}$, and all such Xs bore H3K27me3. About 20% of cells were Xa$^{XIST+}$/Xa$^{XIST-}$ and ~10% were Xi$^{XIST+}$/Xa$^{XIST-}$, and these Xa that lost *XIST* (Xa$^{XIST-}$) still remained H3K27me3$^{+}$. The remaining ~10% were Xi$^{XIST+}$/Xa$^{XIST-}$ with the Xa$^{XIST-}$ with no H3K27me3 (Fig. 1, G and H). Thus, the H3K27me3 domains are the Xs, and the Xa that repressed *XIST* retained H3K27me3 for a certain period. Furthermore, CIZ1, which anchors *XIST* to the X (*32*, *33*), exhibited colocalization with the H3K27me3 domains (fig. S6E), and the TE and ICM had H2AK119u1 domains that were similar to the H3K27me3 domains (fig. S6F). Thus, the Xs express *XIST* and recruit H3K27me3 and H2AK119u1, but these modifications are not sufficient to establish the XCI.

We analyzed the X and the *XIST* coating by three-dimensional structured illumination microscopy (3D-SIM) (*34*). In female CEFs, *XIST* formed a compact cluster of multiple foci with high DAPI density (Xi$^{XIST+}$), whereas *ATRX* exhibited a thread-like signal in the nucleoplasm (Xa$^{XIST-}$) (fig. S7A). In female and male TE at E8 with Xa$^{XIST+}$/Xa$^{XIST+}$ and Xa$^{XIST+}$, respectively, *XIST* exhibited scattered distribution in the nucleoplasm, and *ATRX* showed punctate signals within the *XIST*-expressing areas (Fig. 1I and fig. S7C). On the other hand, female Xi$^{XIST+}$/Xa$^{XIST+}$ TE cells (~2%; Fig. 1B) exhibited one *XIST* domain similar to the Xi$^{XIST+}$ in CEFs and another similar to the Xa$^{XIST+}$ in Xa$^{XIST+}$/Xa$^{XIST+}$ TE, and female Xi$^{XIST+}$/Xa$^{XIST-}$ TE cells showed *XIST* and *ATRX* expression patterns similar to female CEFs (fig. S7C). We quantified the *XIST* dispersion area and the X chromosome territory. The *XIST* dispersion relative to the nuclear size was larger in TE than in CEFs, and it appeared to be larger at E8 than at E9 (Fig. 1, J and K, and fig. S7, B and D); the latter notion was supported by the finding that in female Xa$^{XIST+}$/Xa$^{XIST+}$ cells and male Xa$^{XIST+}$ cells, the *XIST* dispersion areas themselves became smaller from E8 to E9 (Fig. 1M). The *XIST* dispersion areas relative to the X chromosome territories were on average constant (~1) in the TE at E8–E9 and in CEFs (Fig. 1L). Together these results show that the Xa$^{XIST+}$ undergoes a "compaction" but remains active, which is precursory to the XCI establishment or reversion to the Xa$^{XIST-}$ state.

### Differential XCI establishment in key lineages in post-implantation embryos

To elucidate the cascade leading to XCI and to determine the timing of XCI, we analyzed post-implantation embryos. In female cytotrophoblasts (CTs) derived from TE, at E11 (~2 days after implantation), nearly all the cells repressed *XIST* on one X, a majority of them were Xi$^{XIST+}$/Xa$^{XIST-}$, and the Xi$^{XIST+}$ cells bore a condensed morphology with high DAPI density (Fig. 2, A and B, and figs. S8 and S9), indicating that the XCI establishment in CTs occurs rapidly upon implantation. Nonetheless, we noted a substantial heterogeneity in X-linked gene silencing: At E11, *ATRX* underwent inactivation in more than 95% of the cells, whereas *POLA1/HUWE1* remained doubly active in ~40% of the cells (Fig. 2B). We identified ~20% of the cells at E11 as Xa$^{XIST-}$/Xa$^{XIST-}$ for *POLA1/HUWE1* (Fig. 2B). By E17, nearly all the CTs became Xi$^{XIST+}$/Xa$^{XIST-}$ for *HUWE1/ATRX*, whereas more than ~20% remained doubly active for *POLA1* (Fig. 2B).

In the female EPI, at E11, up to ~40% of the cells expressed *XIST* from both Xs, and 30 to 80% were Xa$^{XIST+}$/Xa$^{XIST-}$ (Fig. 2, A and B, and figs. S8 and S9), indicating that the establishment of XCI is protracted in EPI. *POLA1* and *HUWE1* continued to be expressed from both Xs by E13, whereas *ATRX* was repressed in one X more rapidly. Nearly all the EPI exhibited the Xa$^{XIST-}$/Xi$^{XIST+}$ state by E17 (Fig. 2, A and B, and figs. S8 and S9). The Xa$^{XIST+}$ in all stages was characterized by high DAPI intensity, and the Xi$^{XIST+}$ acquired a condensed morphology. In the female amnion (AM) and hypoblast (HYP) derived from ICM, XCI progressed as in the EPI, with slightly faster or slower kinetics in the AM and HYP, respectively, and it was completed by E17 (Fig. 2, A and B, and figs. S8 and S9). A minority (~10%) of cells showed some X-linked gene silencing prior to *XIST* repression (Fig. 2B). In males, all CT, EPI, AM, and HYP showed the Xa$^{XIST-}$ state at E11–E12 (fig. S10, A and B).

Next, we explored the H3K27me3/H2AK119u1 distributions. Until E12, in females, around half of EPI or HYP cells exhibited two H3K27me3/H2AK119u1-enriched domain(s), and the other half showed one H3K27me3/H2AK119u1-enriched domain, which colocalized with DAPI-dense areas (Fig. 2, C to F, and fig. S11). RNA-FISH followed by immunofluorescence on E11 EPI verified that the H3K27me3 domains were the Xs, which were either Xa$^{XIST+}$ (~40%), Xa$^{XIST-}$ (~20%), or Xi$^{XIST+}$ (~40%) (fig. S11B). After E13, more than 80% of EPI and HYP cells bore one H3K27me3/H2AK119u1-enriched domain with high DAPI density (Fig. 2, C to F, and fig. S11). RING1B, a key component of PRC1 (*35*), colocalized with the H2AK119u1-enriched domains at least until E13 and was dispersed thereafter, although the H2AK119u1-enriched domains persisted (Fig. 2, E and F, and fig. S11B). We rarely detected H3K27me3 enrichment in male embryos after E11 (fig. S10B). Collectively, these findings define the timing and differential onset of XCI in four key cell lineages during cy development: In most cells, X chromosome activities proceed from Xa$^{XIST+}$/Xa$^{XIST+}$ to Xa$^{XIST-}$/Xa$^{XIST+}$ (*XIST* down-regulation occurs first), and then to Xa$^{XIST-}$/Xi$^{XIST+}$ (X-linked gene repression follows, with gene-dependent repression timing). To become future Xa, the Xa$^{XIST+}$ that enriches H3K27me3/H2AK119u1 first represses *XIST* and then erases H3K27me3/H2AK119u1, whereas to become future Xi, it continues to express *XIST* and enrich H3K27me3/H2AK119u1, establishing the XCI.

### X:autosome dosage compensation with XCU

We analyzed the allele-specific expression of *XIST* and X-linked genes by single-cell mRNA 3′ end sequencing (SC3-seq) (*31*, *36*, *37*) (table S1 and figs. S12 and S13). The *XIST* expression levels were high and relatively constant from E6 to E16, even though, at E6 to E8, *XIST* was expressed from both Xs, whereas at E13 and E16, it was expressed only from Xp or Xm (Fig. 3, A and B). At E6 to E9, similar to autosomal genes, the X-linked genes were expressed from both alleles, whereas at E13, the X-linked gene ratio with monoallelic expression was increased, and at E16, this trend was strengthened, particularly in the embryonic lineages (Fig. 3C). Consistent with the RNA-FISH analyses, many genes located distally on the short arm (p1), including *POLA1*, continued to show biallelic expression even at E16 (Fig. 3D).

We next investigated the dynamics of the global expression levels of X-linked genes normalized by those of autosomal genes (the X:A ratio). We considered genes with expression levels in the 75th percentile as representative (687 X-linked genes and 16,582 autosomal genes) because most, if not all, housekeeping genes and key genes characterizing each lineage, which are relevant targets for dosage compensation, are those expressed at high levels (*5*) (fig. S14). In the embryonic lineages of males, the X:A ratio showed a progressive, monotonous increase during development, doubling from E6 to E20 (Fig. 3E). In females, the X:A ratio was more constant, but it decreased initially and then regained its original value at E16 (Fig. 3E). The X:A ratio dynamics during the HYP- and TE-lineage development showed similar characteristics (Fig. 3E and fig. S15). The X:A ratio of male cyESCs (CMK6) was similar to those of male cells at E17 and E20, and that of female cyESCs (CMK9), which bore one Xa and one partially eroded X (Xe) (fig. S16) (*19*, *27*, *28*), was higher than that of female cells at E16 (Fig. 4D). Collectively, these findings provide evidence for XCU in males and protracted XCI establishment followed by or coupled with XCU in females.

### X chromosome reactivation in germ cells

Primordial germ cells (PGCs) undergo X chromosome reactivation (XCR) in females. The manner and the mechanism of XCR in PGCs have been unclear, particularly in primates, although it has been reported that in

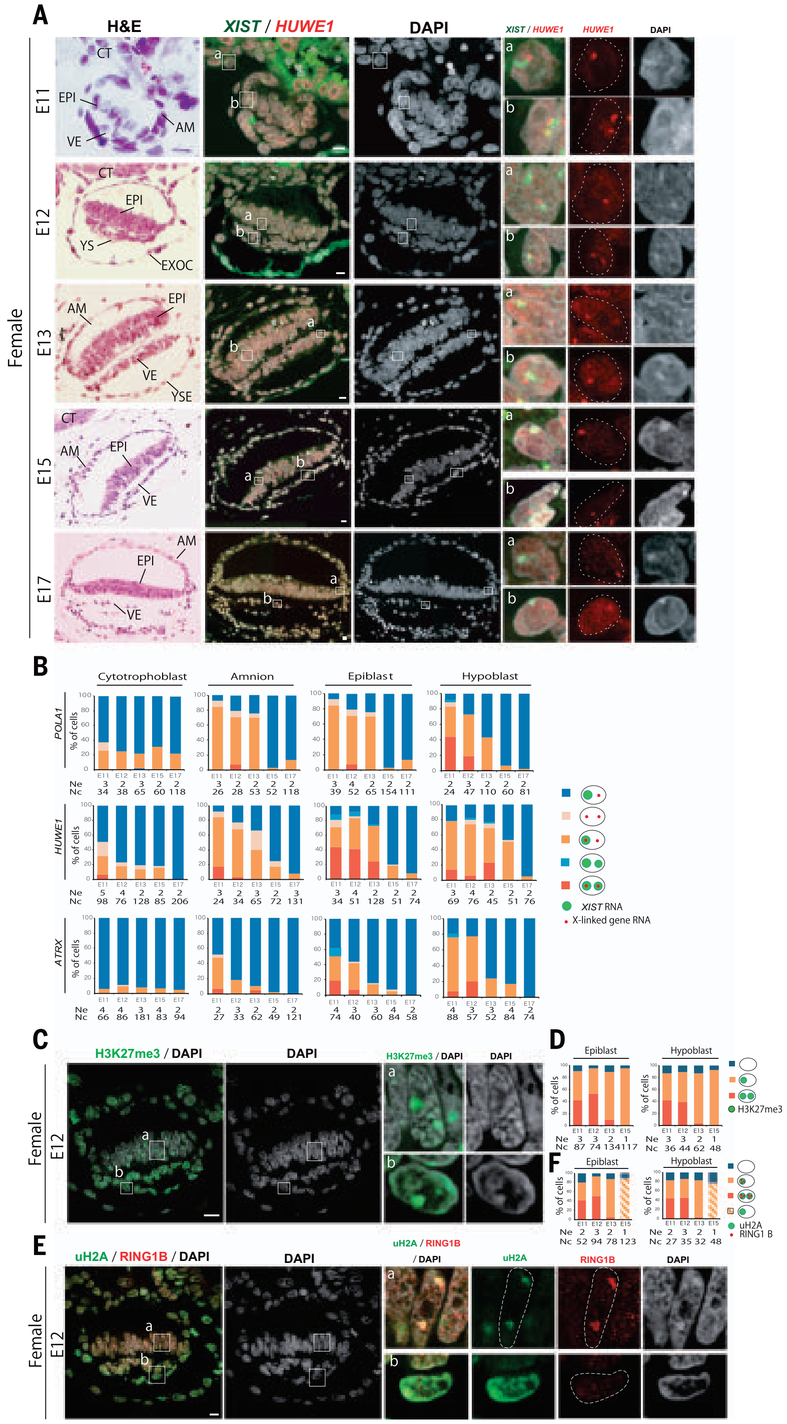

A
H&E
XIST / HUWE1
DAPI
XIST / HUWE1
HUWE1
DAPI
Female
E11
E12
E13
E15
E17
CT
EPI
AM
VE
YS
EXOC
YSE
a
b
B
Cytotrophoblast
Amnion
Epiblast
Hypoblast
POLA1
HUWE1
ATRX
% of cells
E11 E12 E13 E15 E17
Ne 3 2 3 2 2
Nc 34 38 65 60 118
Ne 3 2 2 2 2
Nc 26 28 53 52 118
Ne 3 4 2 2 2
Nc 39 52 65 154 111
Ne 2 3 2 2 2
Nc 24 47 110 60 81
Ne 5 4 2 2 2
Nc 98 76 128 85 206
Ne 3 2 3 2 3
Nc 24 34 65 72 131
Ne 3 3 2 2 2
Nc 34 51 128 51 74
Ne 3 4 2 2 2
Nc 69 76 45 51 76
Ne 4 4 3 4 2
Nc 66 86 181 83 94
Ne 2 3 2 2 2
Nc 27 33 62 49 121
Ne 4 3 3 4 2
Nc 74 40 60 84 58
Ne 4 3 3 4 2
Nc 88 57 52 84 74
XIST RNA
X-linked gene RNA
C
H3K27me3 / DAPI
DAPI
Female E12
D
Epiblast
Hypoblast
Ne 3 3 2 1
Nc 87 74 134 117
Ne 3 3 2 1
Nc 36 44 62 48
H3K27me3
E
uH2A / RING1B / DAPI
DAPI
uH2A
RING1B
F
Ne 2 3 2 1
Nc 52 94 78 123
Ne 2 3 2 1
Nc 27 35 32 48
uH2A
RING1 B

**Fig. 2. X chromosome properties in female cy post-implantation embryos.** (**A**) Sections of female embryos (E11–E17) analyzed by RNA-FISH for *XIST* (green) and *HUWE1* (red) and counterstained by DAPI (gray) and hematoxylin and eosin (H&E). Two representative nuclei (boxed) are magnified. AM, amnion; CT, cytotrophoblast; EPI, epiblast; EXOC, exo-coelomic membrane; VE, visceral endoderm; YS, yolk sac; YSE, yolk-sac endoderm. (**B**) Percentages of female cells (E11–E17) showing distinct *XIST*/X-linked gene expression as indicated by schemes and color codes. (**C**) A section of a female E12 embryo with immunofluorescence for H3K27me3 (green). Two representative nuclei (boxed) (a, EPI; b, YSE) are magnified. (**D**) Percentages of EPI and HYP cells (E11–E15; females) showing distinct H3K27me3 enrichments as indicated by schemes and color codes. (**E**) A section of a female E12 embryo with immunofluorescence for H2AK119u1 (green) and RING1B (red). Two representative nuclei (boxed) (a, EPI; b, YSE) are magnified. (**F**) Percentages of EPI and HYP cells (E11–E15; females) showing distinct H2AK119u1/RING1B enrichments as indicated by schemes and color codes. Scale bars, 10 μm [(A), (C), and (E)].

humans, XCR begins at ~4 weeks of development and *XIST* is activated both in oogonia and gonocytes (*26*, *38*, *39*).

cyPGCs originate in the nascent AM as SOX17$^+$/TFAP2C$^+$ cells at E11 (they may continue to arise until ~E17 from the AM/EPI junction), migrate posteriorly down to the yolk-sac endoderm, and undergo migration through the developing hindgut/mesentery, eventually colonizing the embryonic gonads (*37*). We examined the dynamics of X chromosome activities during cy germ cell development. In females at E13, we detected SOX17$^+$/TFAP2C$^+$ cyPGCs around the AM/EPI junction or the HYP, and they exhibited variegated *XIST* and X-linked gene expression: Like the AM and EPI, ~30 to 60% of cyPGCs expressed *POLA1/HUWE1* biallelically, and less than 20% of cyPGCs expressed *ATRX* biallelically (Fig. 4, A and B, and fig. S17A). On the other hand, unlike the AM and EPI, a larger fraction of cyPGCs (~10 to 40%) repressed *XIST* on both alleles. After E15, in contrast to the AM and EPI, the cyPGC ratio with biallelic X-linked gene expression increased progressively (~60% at E17), with a majority of cyPGCs repressing *XIST* on both Xs (Fig. 4, A and B, and fig. S17A). Accordingly, the cyPGC ratio with XCI decreased progressively from E13 to E17 (Fig. 4B). Thus, cyPGCs initiate XCR upon their specification and their variegated XCI/XCR state may reflect their ongoing specification from the AM/EPI junction.

At E23 and E28 in females, all TFAP2C$^+$ cyPGCs expressed *HUWE1* biallelically, with no or low-level expression of monoallelic *XIST* (Fig. 4, C and D). At E43 and E50 in females, TFAP2C$^+$ cy germ cells (oogonia) expressed *HUWE1* biallelically, with ~70% reexpressing *XIST* from both Xs (Fig. 4D and fig. S17B). In males, all cyPGCs at E28 expressed *HUWE1* but not *XIST*, and a majority of cy germ cells (gonocytes) at E42 expressed *HUWE1*, with reexpression of *XIST* (fig. S18). In females at E23, cyPGCs exhibited either large, single H3K27me3-enriched domains with high DAPI density, most likely the Xs (~40%), or similar domains with smaller size (~60%) (Fig. 4, E and F, and fig. S17C). At E32, ~70% of cyPGCs/oogonia lacked H3K27me3-enriched domains, and at E48, all the oogonia lost such enrichment (Fig. 4, E and F, and fig. S17C). 3D-SIM revealed that in TFAP2C$^+$ oogonia at E57, *XIST* exhibited a highly scattered distribution in two broad areas, with punctate *HUWE1* signals and no DAPI enrichment (Fig. 4G). Early female cyPGCs (~E17) expressed X-linked genes either mono- or biallelically, whereas cy oogonia (~E51) expressed them biallelically (Fig. 4H). The X:A ratio of cy germ cells was increased during their development, with cy oogonia exhibiting an X:A ratio of ~1 (Fig. 3E). Thus, for XCR, cy germ cells first repress *XIST* on both Xs (E13–E17), then erase H3K27me3

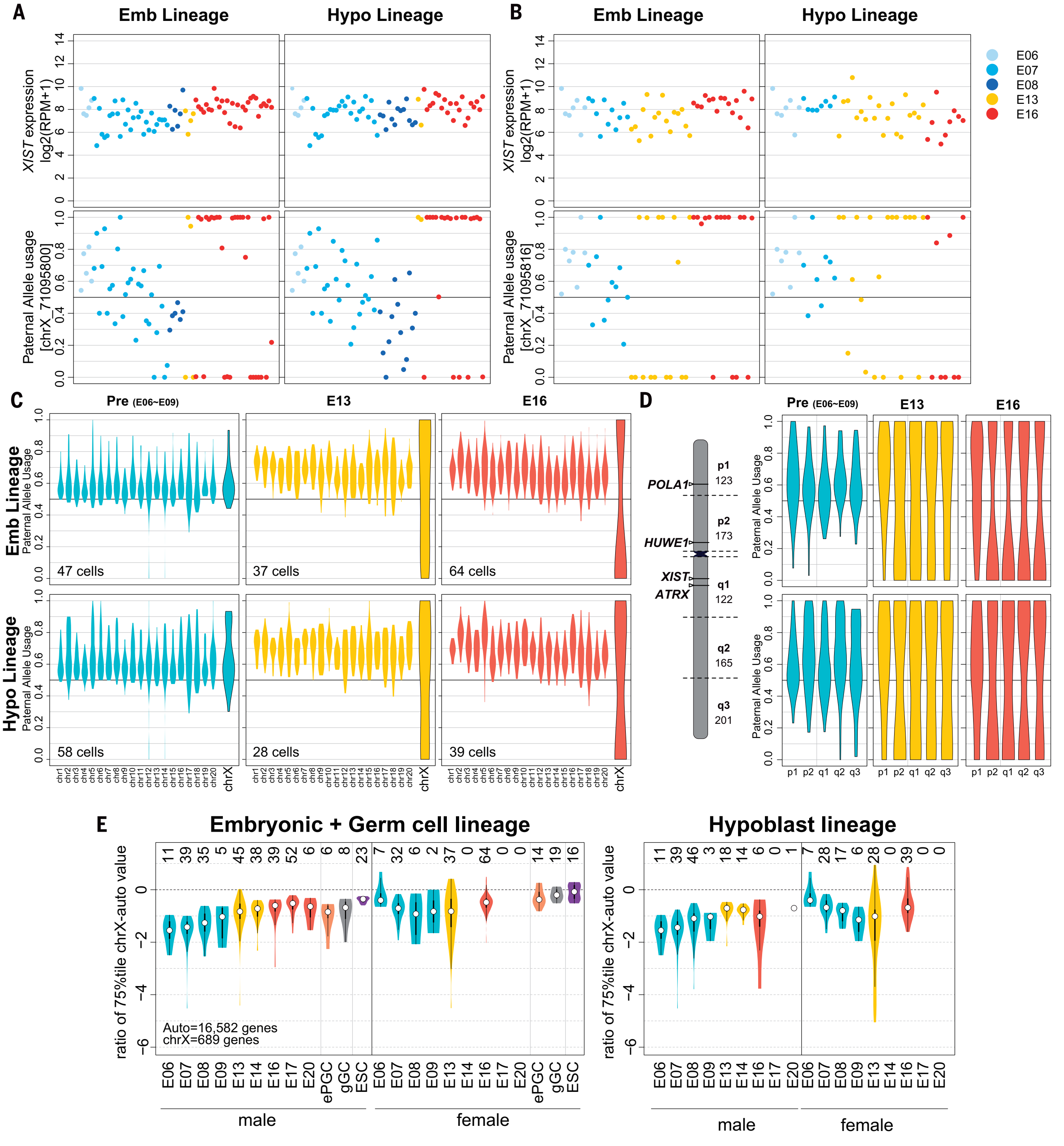


**Fig. 3. SC3-seq for *XIST* and X-linked genes during cy development.** (**A** and **B**) *XIST* expression levels [$\log_2$(RPM + 1)] (top) and ratio [paternal: maternal allele; different polymorphic sites in (A) (chrX_71095800) and (B) (chrX_71095816)] (bottom) measured by SC3-seq in single cells [left, embryonic lineage including ICM, EPI, and GAST; right, hypoblast lineage including ICM, HYP, VEYE, and EXMC (*31*)] at E6, E7, E8, E13, and E16. Each dot represents a value from a single cell, and the color code is as indicated. ICM, inner cell mass; EPI, epiblast; GAST, gastrulating cells; HYP, hypoblast; VEYE, visceral endoderm/yolk-sac endoderm; EXMC, extraembryonic mesenchyme. (**C** and **D**) Median-ratio distributions of the paternal allele usage for autosomal (chromosomes 1 to 20) and X-linked gene expression (C) or X-linked gene expression classified into five groups according to their locations on the X (D, schematic at left) in single cells at E6–E9 (left), E13 (center), and E16 (right) in the embryonic (top) or hypoblast (bottom) lineages. The cell numbers for each category are indicated (B). (**E**) X:A ratio transitions in gene expression in the embryonic and hypoblast lineages in males and females. The genes with the 75th-percentile expression levels in each single cell, among the 16,582 autosomal and 687 X-linked genes in all single cells, were used. The cell numbers used are indicated.

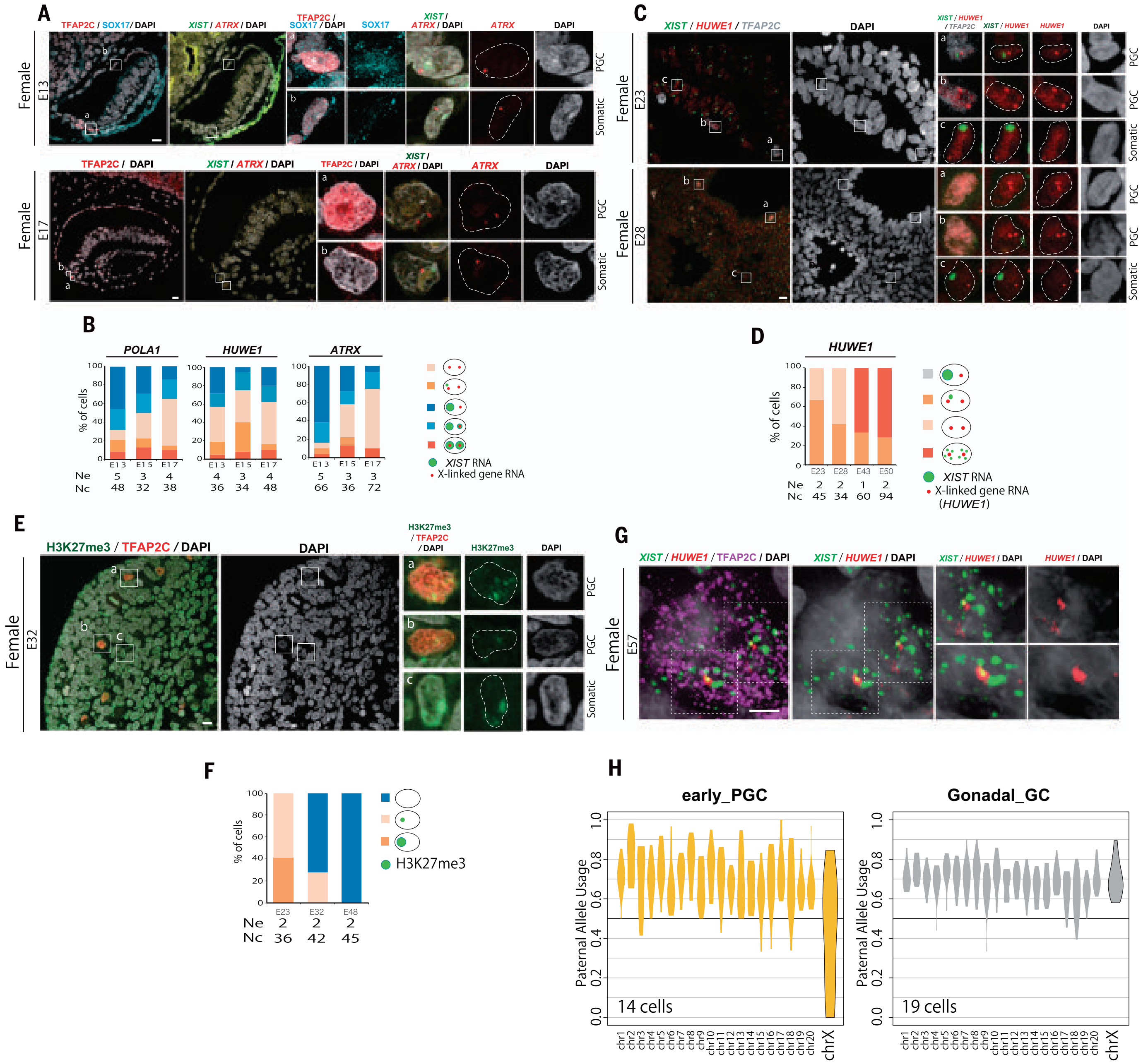

A
TFAP2C / SOX17 / DAPI
XIST / ATRX / DAPI
TFAP2C / SOX17 / DAPI
SOX17
XIST / ATRX / DAPI
ATRX
DAPI
Female E13
Female E17
PGC
Somatic
a
b
TFAP2C / DAPI
B
POLA1
HUWE1
ATRX
% of cells
E13 E15 E17
Ne 5 3 4
Nc 48 32 38
Ne 4 3 4
Nc 36 34 48
Ne 5 3 3
Nc 66 36 72
XIST RNA
X-linked gene RNA
C
XIST / HUWE1 / TFAP2C
DAPI
XIST / HUWE1
HUWE1
Female E23
Female E28
D
HUWE1
E23 E28 E43 E50
Ne 2 2 1 2
Nc 45 34 60 94
XIST RNA
X-linked gene RNA (HUWE1)
E
H3K27me3 / TFAP2C / DAPI
H3K27me3
Female E32
c
F
E23 E32 E48
Ne 2 2 2
Nc 36 42 45
H3K27me3
G
XIST / HUWE1 / TFAP2C / DAPI
XIST / HUWE1 / DAPI
HUWE1 / DAPI
Female E57
H
early_PGC
Gonadal_GC
Paternal Allele Usage
14 cells
19 cells
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chrX
0.0 0.2 0.4 0.6 0.8 1.0

**Fig. 4. X chromosome properties in female cy germ cells.** (**A**) Sections of female E13 and E17 embryos analyzed by RNA-FISH for *XIST* (green) and *ATRX* (red) with immunofluorescence for TFAP2C (red) and SOX17 (cyan). DAPI is shown in gray. For each day, two representative nuclei (boxed) (a, PGC; b, soma) are magnified. (**B**) Percentages of female PGCs (E11–E15) analyzed as in (A) showing distinct *XIST*/X-linked gene expression as indicated by schemes and color codes. (**C**) Sections containing female PGCs (E23–E28) analyzed by RNA-FISH for *XIST* (green) and *HUWE1* (red) with immunofluorescence for TFAP2C (cyan/gray). DAPI (gray). For each day, three representative nuclei (boxed) (a and b, PGCs; c, soma) are magnified. (**D**) Percentages of female PGCs (E23–E50) analyzed as in (C) showing distinct *XIST*/*HUWE1* expression as indicated by schemes and color codes. (**E**) A section of an incipient female E32 gonad containing germ cells analyzed by immunofluorescence for H3K27me3 (green) and TFAP2C (red). Three representative nuclei (boxed) (a and b, germ cells; c, soma) are magnified. (**F**) Percentage of female germ cells at E23, E32, and E48 showing distinct H3K27me3 enrichments as indicated by schemes and color codes. (**G**) 3D-SIM image of *XIST* RNA (green), *HUWE1* (red), and TFAP2C (purple) in E57 female germ cells. Two X chromosomes (dotted boxes) are magnified. (**H**) Median-ratio distributions of the paternal allele usage for autosomal (chromosomes 1 to 20) and X-linked gene expression in single cy germ cells measured at E13, E16, and E17 (left) or E43, E50, and E51 (right). The cell numbers analyzed are as indicated. Scale bars, 10 μm [(A), (C), and (E)], 2 μm (G).

(E22–E28) and reexpress *XIST* thereafter (E42). The X-linked genes show biallelic expression without repression or with reactivation at an early stage (E13) and thereafter gain a full biallelic expression.

## Discussion

We have delineated the X chromosome dosage compensation program during cy development (Fig. 5 and fig. S19): Both Xp and Xm express *XIST* and recruit H3K27me3 and H2AK119u1. The Xs with this state show a progressive "compaction" but remain active. Future Xa represses *XIST* and erases repressive modifications with progressive XCU, whereas future Xi maintains *XIST* and repressive

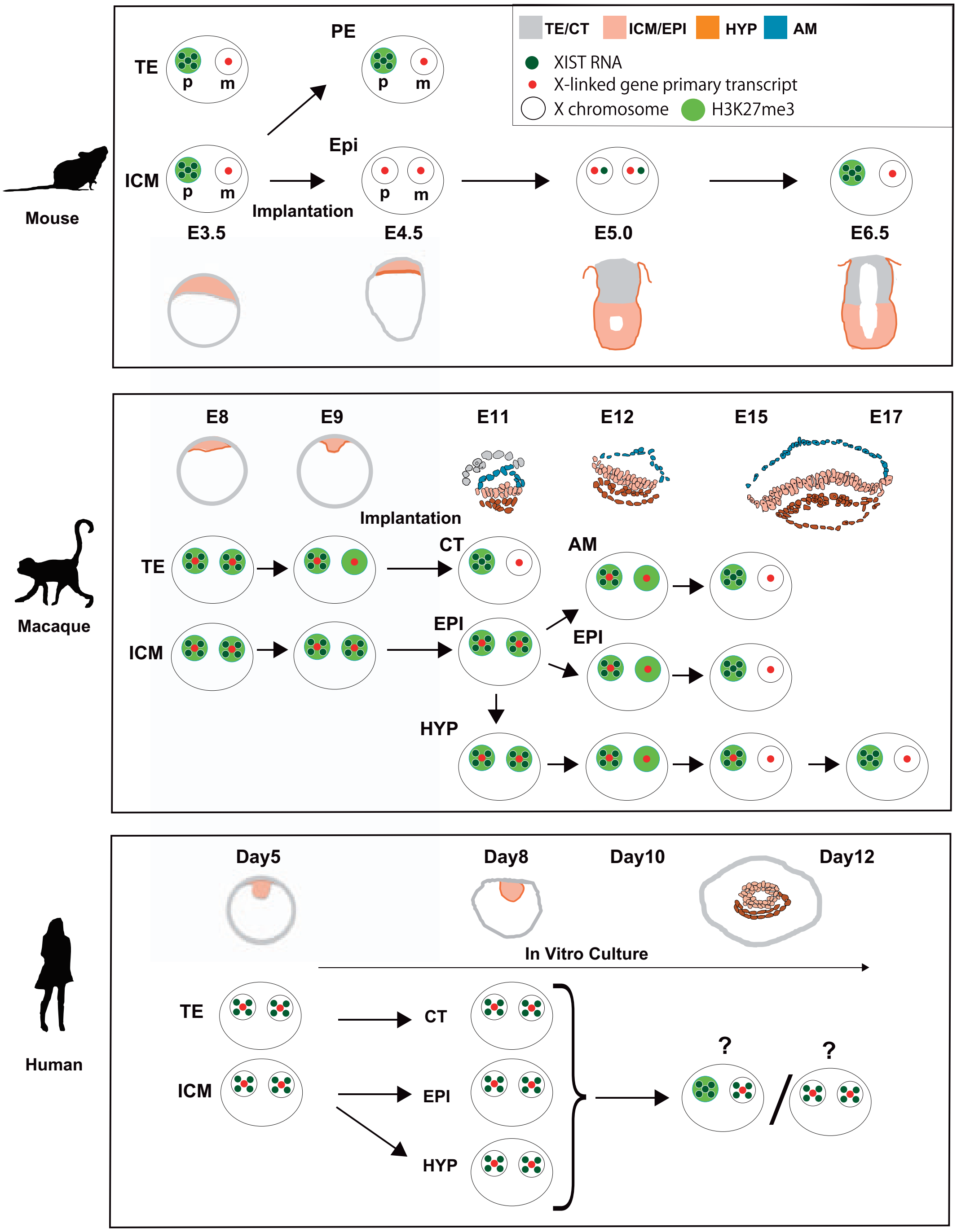

TE/CT
ICM/EPI
HYP
AM
XIST RNA
X-linked gene primary transcript
X chromosome
H3K27me3
Mouse
TE
PE
ICM
Epi
p
m
Implantation
E3.5
E4.5
E5.0
E6.5
Macaque
E8
E9
E11
E12
E15
E17
Implantation
TE
ICM
CT
EPI
AM
EPI
HYP
Human
Day5
Day8
Day10
Day12
In Vitro Culture
TE
ICM
CT
EPI
HYP
?
?

**Fig. 5. A model for X chromosome dynamics during development in mice, monkeys, and humans.** The schematic summarizes X chromosome dynamics during development in mice (6, 13, 14, 16, 40), monkeys (this study), and humans (20, 25, 26). Note that the model in humans is based on in vitro culture studies.

modifications, and condenses the whole chromosome for XCI. These sequences of events occur in the embryonic and extraembryonic lineages and in both males and females, with earlier and protracted XCI establishment in TE and CT in males and in AM, EPI, and HYP in females. Given the imprinted and early XCI in HYP and visceral endoderm in mice (40), the protracted XCI establishment in HYP was unexpected, suggesting a mechanism in which XCI is initiated independently of the pluripotency circuitry in primates.

H3K27me3 and H2AK119u1 accumulate in cy $\mathrm{Xa}^{XIST+}$, and such a state persists after implantation until E12. This accumulation does not occur in $\mathrm{Xa}^{XIST+}$ in human pre-implantation embryos; this has been regarded as an explanation for why human $\mathrm{Xa}^{XIST+}$ is not immediately subject to repression (20). Although the case in humans will require further examination, the above findings may reflect the divergence of mechanisms that recruit or repel *XIST* to or from the X among primates. For example, in human pluripotent cells, an X-linked lncRNA, *XACT*, coats Xa and has been proposed to compete with *XIST* and to prevent the repressive activity of *XIST* (23). *XACT* is specific to hominids and not conserved in macaques (41), potentially explaining why cy $\mathrm{Xa}^{XIST+}$, but not human $\mathrm{Xa}^{XIST+}$, accumulates H3K27me3 and H2AK119u1. The mechanism by which the X remains active with repressive modifications is unknown. An investigation into the chromatin landscape of the X during cy development would provide a clue to understanding this mechanism. Our findings also suggest an undefined mechanism that is essential for triggering chromosome-wide XCI.

There has been conflicting evidence as to the transcriptional state of human Xa$^{XIST+}$. One study showed that although Xa$^{XIST+}$ is still active, its overall expression is dampened, thereby establishing XCI (*21*), whereas work based on the same or similar datasets has provided no such evidence (*22*, *23*). Our analysis involving the genes with expression levels in the 75th percentile, which represent genes relevant to dosage compensation, revealed that in cynomolgus monkeys, there might exist a dampening-like event; if so, however, it operates transiently, does not ensure the male (single X) level, and is not manifest in males bearing Xa$^{XIST+}$ with H3K27me3, which suggests that it may not be a major mechanism for dosage compensation. The dynamics of total X-linked gene expression, calculated in a manner similar to that of Petropoulos *et al.* (*21*), can be seen in fig. S20. The mechanism of XCU requires further investigation.

We have clarified a cascade of events for XCR in cy germ cells, in which *XIST* repression and the ensuing erasure of repressive histone modifications occur by E28 during their migration (Fig. 3 and fig. S19). Cy germ cells colonized in the gonads reactivate *XIST* both in females and males by an unknown mechanism, yet *XIST* coats the Xs in a highly scattered manner, without H3K27me3 recruitment. Thus, cy germ cells bear an Xa$^{XIST+}$ distinct from those in other early embryonic cells, which might represent a noncanonical X dosage compensation state (*42*). Gonadal germ cells in mice do not reactivate *Xist* despite their naïve epigenome after epigenetic reprogramming (*43*, *44*). Thus, *XIST* regulation in germ cells involves an evolutionarily divergent mechanism between primates and mice. Our study clarifies that the ontogeny and distinct mechanism of X chromosome dosage compensation are critically divergent in cynomolgus monkeys and mice; this is directly relevant to female pluripotent stem cell biology in primates, including humans, in which an appropriate regulation of the X chromosome activities is a critical requirement.

## Methods summary

For the generation of cy pre-implantation embryos, MII oocytes collected by laparoscopy were subjected to intracytoplasmic sperm injection (ICSI) and embryo culture. For the generation of post-implantation embryos, cultured pre-implantation embryos were transferred into female recipients. Embryo implantations were monitored by transabdominal ultrasound scanning, and embryos were recovered either by hysterectomy (E11–E17) or cesarean section (E20–) under full anesthesia. For immunofluorescence, RNA-FISH, and DNA-FISH analyses of pre-implantation embryos, all procedures were carried out with the embryos attached to coverslips coated with Denhardt's solution. For the immunofluorescence after RNA-FISH of pre-implantation embryos, embryos attached on coverslips were incubated in PBS containing proteinase K (10 μg/ml) at 37°C for 15 min and subjected to the immunofluorescence analysis. Immunofluorescence and RNA-FISH of early post-implantation embryos (E11–E17) or late post-implantation embryos (E20–) were performed with formalin-fixed paraffin-embedded (FFPE) embryo sections or frozen embryo sections on the slide. For single-cell RNA-seq analysis, the cDNA libraries and the sequence data were generated with the SC3-seq method (*31*, *36*, *37*). For the analysis of XCI/XCU, the read count data were normalized to reads per million mapped reads (RPM) and transformed into $\log_2$ scale [$\log_2(\mathrm{RPM} + 1)$]. The X:A ratio was calculated by dividing the 75th percentile $\log_2(\mathrm{RPM} + 1)$ values of the X-linked genes by those of the autosomal genes in each cell. For the allele usage analysis, pseudo-embryonic genomes were generated by combining the SC3-seq data in each embryo. Then the paternal allele on the single-nucleotide polymorphism positions were called by genome information of sperm donor and gatk-HaplotypeCaller. All the sequence data in this study are available (see the acknowledgments).

## REFERENCES AND NOTES

1. S. Ohno, *Sex Chromosomes and Sex-Linked Genes* (Springer, 1967).
2. M. T. Ross *et al*., The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005). doi: 10.1038/nature03440; pmid: 15772651
3. H. Skaletsky *et al*., The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003). doi: 10.1038/nature01722; pmid: 12815422
4. M. F. Lyon, Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* **190**, 372–373 (1961). doi: 10.1038/190372a0; pmid: 13764598
5. X. Deng, J. B. Berletch, D. K. Nguyen, C. M. Disteche, X chromosome regulation: Diverse patterns in development, tissues and disease. *Nat. Rev. Genet.* **15**, 367–378 (2014). doi: 10.1038/nrg3687; pmid: 24733023
6. J. J. Żylicz, E. Heard, Molecular Mechanisms of Facultative Heterochromatin Formation: An X-Chromosome Perspective. *Annu. Rev. Biochem.* **89**, 255–282 (2020). pmid: 32259458
7. N. Takagi, K. Abe, Detrimental effects of two active X chromosomes on early mouse development. *Development* **109**, 189–201 (1990). doi: 10.1242/dev.109.1.189; pmid: 2209464
8. D. M. Snell, J. M. A. Turner, Sex Chromosome Effects on Male-Female Differences in Mammals. *Curr. Biol.* **28**, R1313–R1324 (2018). doi: 10.1016/j.cub.2018.09.018; pmid: 30458153
9. N. Brockdorff *et al*., The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515–526 (1992). doi: 10.1016/0092-8674(92)90519-I; pmid: 1423610
10. G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff, Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996). doi: 10.1038/379131a0; pmid: 8538762
11. Y. Marahrens, B. Panning, J. Dausman, W. Strauss, R. Jaenisch, Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* **11**, 156–166 (1997). doi: 10.1101/gad.11.2.156; pmid: 9009199
12. S. H. Namekawa, B. Payer, K. D. Huynh, R. Jaenisch, J. T. Lee, Two-step imprinted X inactivation: Repeat versus genic silencing in the mouse. *Mol. Cell. Biol.* **30**, 3187–3205 (2010). doi: 10.1128/MCB.00227-10; pmid: 20404085
13. W. Mak *et al*., Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303**, 666–669 (2004). doi: 10.1126/science.1092674; pmid: 14752160
14. I. Okamoto, A. P. Otte, C. D. Allis, D. Reinberg, E. Heard, Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644–649 (2004). doi: 10.1126/science.1092727; pmid: 14671313
15. C. Patrat *et al*., Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5198–5203 (2009). doi: 10.1073/pnas.0810683106; pmid: 19273861
16. N. Takagi, M. Sasaki, Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* **256**, 640–642 (1975). doi: 10.1038/256640a0; pmid: 1152998
17. A. Inoue, L. Jiang, F. Lu, Y. Zhang, Genomic imprinting of *Xist* by maternal H3K27me3. *Genes Dev.* **31**, 1927–1932 (2017). doi: 10.1101/gad.304113.117; pmid: 29089420
18. A. Fukuda *et al*., The role of maternal-specific H3K9me3 modification in establishing imprinted X-chromosome inactivation and embryogenesis in mice. *Nat. Commun.* **5**, 5464 (2014). doi: 10.1038/ncomms6464; pmid: 25394724
19. C. Patrat, J. F. Ouimette, C. Rougeulle, X chromosome inactivation in human development. *Development* **147**, dev183095 (2020). doi: 10.1242/dev.183095; pmid: 31900287
20. I. Okamoto *et al*., Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011). doi: 10.1038/nature09872; pmid: 21471966
21. S. Petropoulos *et al*., Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**, 1012–1026 (2016). doi: 10.1016/j.cell.2016.03.023; pmid: 27062923
22. J. C. Moreira de Mello, G. R. Fernandes, M. D. Vibranovski, L. V. Pereira, Early X chromosome inactivation during human preimplantation development revealed by single-cell RNA-sequencing. *Sci. Rep.* **7**, 10794 (2017). doi: 10.1038/s41598-017-11044-z; pmid: 28883481
23. C. Vallot *et al*., XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. *Cell Stem Cell* **20**, 102–111 (2017). doi: 10.1016/j.stem.2016.10.014; pmid: 27989768
24. Q. Zhou *et al*., Single-cell RNA-seq reveals distinct dynamic behavior of sex chromosomes during early human embryogenesis. *Mol. Reprod. Dev.* **86**, 871–882 (2019). doi: 10.1002/mrd.23162; pmid: 31094050
25. F. Zhou *et al*., Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature* **572**, 660–664 (2019). doi: 10.1038/s41586-019-1500-0; pmid: 31435013
26. T. Chitiashvili *et al*., Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. *Nat. Cell Biol.* **22**, 1436–1446 (2020). doi: 10.1038/s41556-020-00607-4; pmid: 33257808
27. S. Mekhoubad *et al*., Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell* **10**, 595–609 (2012). doi: 10.1016/j.stem.2012.02.014; pmid: 22560080
28. K. L. Nazor *et al*., Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* **10**, 620–634 (2012). doi: 10.1016/j.stem.2012.02.013; pmid: 22560082
29. M. Tachibana *et al*., X-chromosome inactivation in monkey embryos and pluripotent stem cells. *Dev. Biol.* **371**, 146–155 (2012). doi: 10.1016/j.ydbio.2012.08.009; pmid: 22935618
30. X. Wang *et al*., Transcriptome analyses of rhesus monkey preimplantation embryos reveal a reduced capacity for DNA double-strand break repair in primate oocytes and early embryos. *Genome Res.* **27**, 567–579 (2017). doi: 10.1101/gr.198044.115; pmid: 28223401
31. T. Nakamura *et al*., A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57–62 (2016). doi: 10.1038/nature19096; pmid: 27556940

32. R. Ridings-Figueroa *et al*., The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* **31**, 876–888 (2017). doi: 10.1101/gad.295907.117; pmid: 28546514
33. H. Sunwoo, D. Colognori, J. E. Froberg, Y. Jeon, J. T. Lee, Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10654–10659 (2017). doi: 10.1073/pnas.1711206114; pmid: 28923964
34. R. Heintzmann, T. Huser, Super-Resolution Structured Illumination Microscopy. *Chem. Rev.* **117**, 13890–13908 (2017). doi: 10.1021/acs.chemrev.7b00218; pmid: 29125755
35. M. de Napoles *et al*., Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* **7**, 663–676 (2004). doi: 10.1016/j.devcel.2004.10.005; pmid: 15525528
36. T. Nakamura *et al*., SC3-seq: A method for highly parallel and quantitative measurement of single-cell gene expression. *Nucleic Acids Res.* **43**, e60 (2015). doi: 10.1093/nar/gkv134; pmid: 25722368
37. K. Sasaki *et al*., The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. *Dev. Cell* **39**, 169–185 (2016). doi: 10.1016/j.devcel.2016.09.007; pmid: 27720607
38. F. Guo *et al*., The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* **161**, 1437–1452 (2015). doi: 10.1016/j.cell.2015.05.015; pmid: 26046443
39. W. W. Tang *et al*., A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* **161**, 1453–1467 (2015). doi: 10.1016/j.cell.2015.04.053; pmid: 26046444
40. N. Takagi, O. Sugawara, M. Sasaki, Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. *Chromosoma* **85**, 275–286 (1982). doi: 10.1007/BF00294971; pmid: 6180866
41. C. Vallot *et al*., XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.* **45**, 239–241 (2013). doi: 10.1038/ng.2530; pmid: 23334669
42. M. N. Sangrithi *et al*., Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline. *Dev. Cell* **40**, 289–301.e3 (2017). doi: 10.1016/j.devcel.2016.12.023; pmid: 28132849
43. M. Sugimoto, K. Abe, X chromosome reactivation initiates in nascent primordial germ cells in mice. *PLOS Genet.* **3**, e116 (2007). doi: 10.1371/journal.pgen.0030116; pmid: 17676999
44. S. M. Chuva de Sousa Lopes *et al*., X chromosome activity in mouse XX primordial germ cells. *PLOS Genet.* **4**, e30 (2008). doi: 10.1371/journal.pgen.0040030; pmid: 18266475

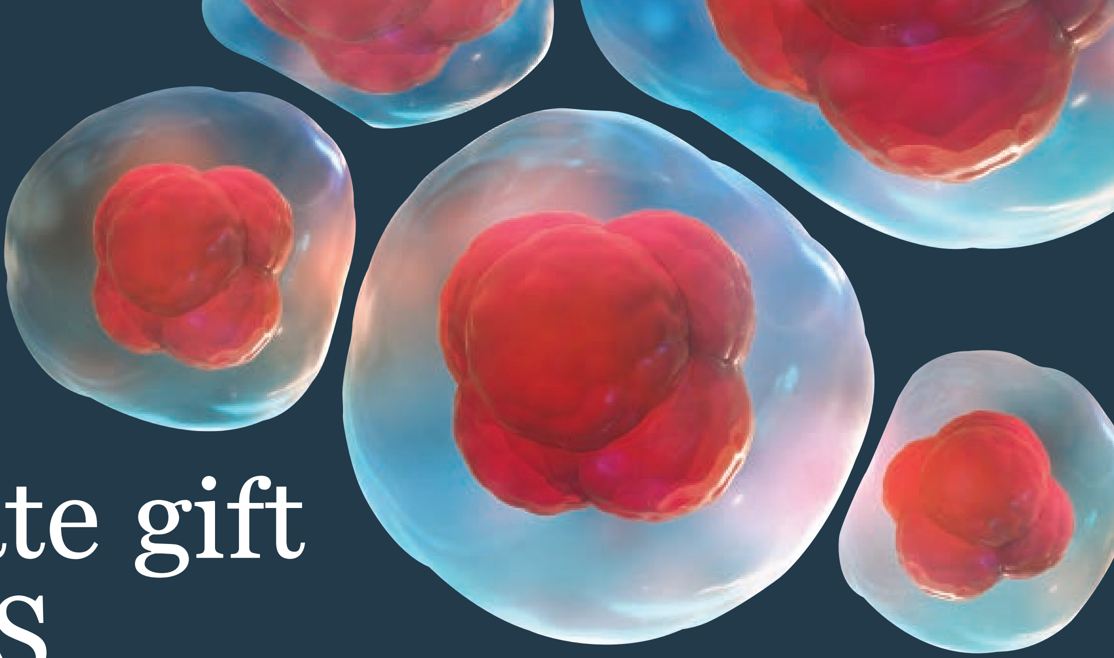
## ACKNOWLEDGMENTS

We thank the members of our laboratory for their helpful input on this study; and Y. Nagai, N. Konishi, E. Tsutsumi, and M. Kawasaki of the Saitou laboratory and T. Sato and M. Kabata of the Yamamoto laboratory for technical assistance; the Center for Anatomical, Pathological and Forensic Medical Research at Kyoto University for the preparation of paraffin sections; the animal care staff at the Research Center, Shiga University of Medical Science, for their assistance with the animals; Nikon Corp. for the superresolution imaging microscopic analysis; and the Single-Cell Genome Information Analysis Core (SignAC) in ASHBi for the RNA sequence analysis. **Author contributions:** I.O. conducted the immunofluorescence/FISH experiments on cy embryos and analyzed the data. T.N. and I.O. performed the SC3-seq experiments, and T.N. analyzed the SC3-seq data. Y.Y. and T.Y. contributed to the SC3-seq data analysis. K.S., C.I., H.T., S.N., and M.E. contributed to the isolation of cy embryos. I.O. and M.S. conceived the project, designed the experiments, and wrote the manuscript. **Funding:** Supported by JSPS Grant-in-Aid for Scientific Research (C) 18K06030 (I.O.); MEXT Grant-in-Aid for Transformative Research Areas B 20H05761 (T.N. and I.O.); and JSPS Grant-in-Aid for Specially Promoted Research 17H06098, JST-ERATO grant JPMJER1104, HFSP grant RGP0057/2018, and grants from the Pythias Fund and Open Philanthropy Project (M.S.). **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. The accession numbers for the SC3-seq and WGS data generated in this study are as follows. For the SC3-seq data of cy embryos: GSE151149 (GEO database). For the WGS data of male monkeys (sperm donors): DRA010279 (DDBJ database).

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abd8887
Materials and Methods
Figs. S1 to S20
Tables S1 and S2
References (*45*–*58*)

18 July 2020; accepted 12 October 2021
10.1126/science.abd8887



1848 SOCIETY
AAAS

# An estate gift to AAAS

**Going all the way back to 1848,** our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

Today, we are dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy and for the community to speak with one voice to advance science and engineering in the United States and around the world.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science and society – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

> *"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."*
>
> **—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member**

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

**Legal Name:** American Association for the Advancement of Science
**Federal Tax ID Number:** 53-0196568
**Address:** 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

**PLEASE CONTACT ME AT:**

1848 SOCIETY
AAAS

Name: ______________________________

Address: ______________________________

City: ______________ State: ________ Zip code: ______________ Country: __________

Email: ______________________ Phone: ______________________

**RETURN THIS FORM TO:**
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLES

VIROLOGY

# A viral RNA hijacks host machinery using dynamic conformational changes of a tRNA-like structure

**Steve L. Bonilla[1], Madeline E. Sherlock[1], Andrea MacFadden[1], Jeffrey S. Kieft[1,2]***

**Viruses require multifunctional structured RNAs to hijack their host's biochemistry, but their mechanisms can be obscured by the difficulty of solving conformationally dynamic RNA structures. Using cryo–electron microscopy (cryo-EM), we visualized the structure of the mysterious viral transfer RNA (tRNA)–like structure (TLS) from the brome mosaic virus, which affects replication, translation, and genome encapsidation. Structures in isolation and those bound to tyrosyl-tRNA synthetase (TyrRS) show that this ~55-kilodalton purported tRNA mimic undergoes large conformational rearrangements to bind TyrRS in a form that differs substantially from that of tRNA. Our study reveals how viral RNAs can use a combination of static and dynamic RNA structures to bind host machinery through highly noncanonical interactions, and we highlight the utility of cryo-EM for visualizing small, conformationally dynamic structured RNAs.**

RNA's functional versatility derives from its ability to encode genetic information and form complex three-dimensional (3D) structures (*1*). RNA viruses exploit these features, using structured RNA elements to manipulate host machinery and regulate essential viral processes (*2*, *3*). Such RNA elements exist in viral clades as diverse as flaviviruses, lentiviruses, coronaviruses, alphaviruses, and picornaviruses (*2*–*7*), where often a single RNA element performs multiple functions. Our understanding of such RNAs is rudimentary, partly because many are conformationally dynamic and therefore difficult to characterize structurally. Crystallization of such elements is difficult because of their dynamic structure, and nuclear magnetic resonance (NMR) is often not tractable for fully functional RNAs or RNA-protein complexes because of molecular weight limitations (*8*, *9*). Thus, structure-function rules of viral RNAs and their interactions with host machinery are not well understood.

A powerful way that viruses use RNA structure is to mimic cellular tRNAs. Several viral internal ribosome entry sites (IRESs), including that of the hepatitis C virus, mimic parts of tRNA to contact tRNA binding sites on the ribosome (*2*). Additionally, the the HIV-1 RNA genome contains a tRNA-like element that binds lysyl-tRNA synthetase and favors the release of bound tRNA$^{Lys3}$, the primer for HIV reverse transcription (*5*). These and other examples show the importance of tRNA mimicry in diverse viruses, including some that cause human diseases.

Important examples of tRNA mimicry and multifunctionality are the tRNA-like structures (TLSs) at the 3′ ends of certain positive-strand RNA viral genomes (*10*, *11*). TLSs drive aminoacylation of viral genomic 3′ ends by host aminoacyl-tRNA synthetases (aaRSs) and can interact with other tRNA-specific enzymes, including CCA-nucleotidyltransferase (CCA-NTase) and eukaryotic translation elongation factor 1A (eEF1A) (*10*, *12*). Known TLSs are classified into three types on the basis of their aaRS specificity: valylatable (TLS$^{Val}$), histidylatable (TLS$^{His}$), and tyrosylatable (TLS$^{Tyr}$) (*10*, *12*). Each type has a characteristic secondary structure that differs substantially from that of tRNA, showing that tRNA mimicry can be achieved in diverse ways (fig. S1) (*12*). Of the three types, TLS$^{Tyr}$ differs most from tRNA in size and secondary structure and is the most difficult to reconcile with tRNA mimicry (*13*). The prototype TLS$^{Tyr}$ is found at the 3′ end of each of the three genomic RNAs of the tripartite brome mosaic virus (BMV) genome (Fig. 1A). The BMV TLS$^{Tyr}$ (hereafter referred to as BMV TLS) plays roles in translation, replication, and encapsidation of BMV RNAs; some of these functions are linked to aminoacylation of the TLS, whereas other functions appear to be independent of aminoacylation status (*10*, *13*). The BMV TLS is thus a powerful model system to explore important viral RNA features, including multifunctionality, tRNA mimicry, host protein binding, and potentially conformational dynamics.

Early studies identified the 3′-most 134 nucleotides (nts) of the genomic BMV RNAs as the minimal tyrosylatable TLS RNA (*13*, *14*), but later studies have demonstrated the importance of the adjacent upstream sequence (*13*). The consensus TLS contains 169 nts that form a secondary structure with seven helical stems (compared with four for tRNA), including a pseudoknotted aminoacyl acceptor stem analog (Fig. 1A and fig. S1) (*12*, *15*). Thus, BMV TLS is substantially larger and more structurally complex than tRNA and is therefore an example of a structure that accomplishes tRNA mimicry in a manner that is not readily apparent. It is not known whether BMV TLS contains a tRNA-like L-shaped structure, and conflicting evidence points to either stem B2 or B3 as the anticodon stem analog (Fig. 1A) (*16*). Notwithstanding the lessons that a 3D structure of the BMV TLS could provide about viral RNA structure-function relationships, the structure has remained elusive for decades.

## Cryo–electron microscopy (cryo-EM) reveals the global architecture of unbound BMV TLS

We transcribed and purified a TLS sequence from BMV RNA 3 in vitro (*12*, *15*) and confirmed its ability to be tyrosylated in vitro using recombinantly expressed tyrosyl-tRNA synthetase (TyrRS) from model host *Phaseolus vulgaris* (Fig. 1B). Cryo-EM micrographs of this sample collected using a 200-kV microscope contained readily identifiable particles despite their small size (55 kDa) (Fig. 1C and fig. S2). Using a data analysis pipeline that includes the removal of junk particles through 2D classification, ab initio 3D reconstruction, 3D classification, and map refinement (*17*), we obtained an initial 7.0-Å map that displayed characteristics consistent with those of a folded RNA of the expected size (Fig. 1D and fig. S2). The overall architecture of the map was robustly reproduced across ab initio reconstructions with different numbers of classes (fig. S2).

A larger dataset from a 300-kV microscope equipped with an energy filter increased the overall resolution to 4.3 Å. In this map, both minor and major grooves of A-form helices were clearly defined, and the connectivity of the phosphate backbone could be deduced (Fig. 1E and fig. S3). In some regions, stacking and coplanarity of base pairs were resolved, and phosphate bumps were visible (fig. S4). The central core of the map displayed the highest local resolution, but even peripheral-region density was well defined and displayed clear helical features (Fig. 1E).

Notably, one helical domain stood out because it was less defined and had lower local resolution relative to that of other map regions (Fig. 1E, boxed), which suggests local flexibility within the structure. To examine this, we used 3D variability analysis (*18*) to generate a series of 3D volumes representing variability among particles within the dataset (Fig. 1F). This analysis was consistent with this one helical domain occupying multiple conformational states while the rest of the RNA is relatively more static.

---

[1] Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA. [2] RNA BioScience Initiative, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA.
***Corresponding author. Email: jeffrey.kieft@cuanschutz.edu**

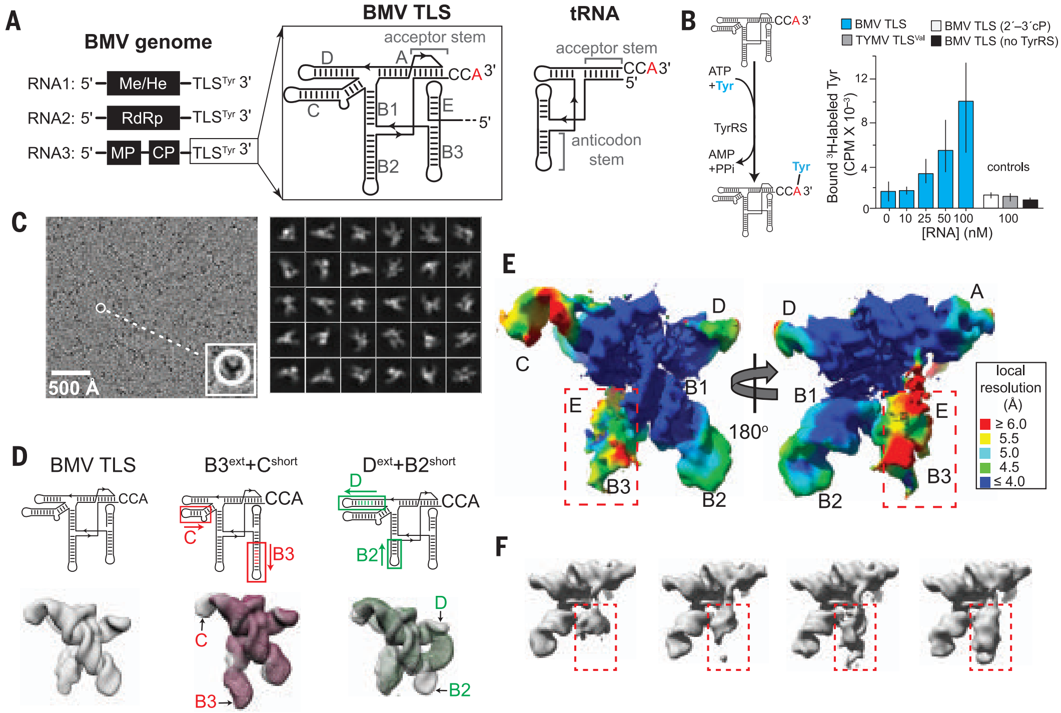


**Fig. 1. Functional and initial structural characterization of BMV TLS RNA.** (**A**) Organization of the tripartite BMV genome with a TLS at the 3′ end of each RNA. The conserved BMV TLS secondary structure is shown next to that of a tRNA. Structural domains and the terminal CCA are shown, with the red "A" designating the aminoacylation site. Me/He, methyltransferase/helicase; RdRp, RNA-dependent RNA polymerase; MP, movement protein; CP, coat protein. (**B**) (Left) Cartoon of reaction catalyzed by TyrRS. ATP, adenosine 5′-triphosphate; AMP, adenosine 5′-monophosphate. (Right) Tyrosylation of the BMV TLS RNA used for cryo-EM, using an adapted published protocol (*36*). BMV TLS (2′–3′ cP) contains a terminal 2′–3′ cyclic phosphate, not an efficient substrate of TyrRS. TYMV $TLS^{Val}$ is a valylatable TLS from turnip yellow mosaic virus (TYMV). CPM, counts per minute. (**C**) (Left) Representative micrograph of the BMV TLS RNA, 30% of the field of view. Defocus range was −1 to −2.5 μm. (Right) Classified projections from BMV TLS particles (fig. S2). (**D**) Secondary structures and cryo-EM maps of BMV TLS, $B3^{ext}$ + $C^{short,}$ and $D^{ext}$ + $B2^{short}$. The maps of $B3^{ext}$ + $C^{short}$ (mesh red) and $D^{ext}$ + $B2^{short}$ (mesh green) are superimposed on BMV TLS (solid gray). Arrows point to differences corresponding to altered stems. (**E**) Refined map of BMV TLS at 4.3-Å resolution. Colors denote local resolution. The flexible domain with lower local resolution is boxed in both views. (**F**) Maps representing heterogeneity in the particles using 3D variability analysis (*18*). The variability is mostly localized to the domain that is boxed in red in each map.

### Engineered RNAs provide information for unambiguous structural modeling

Aside from low-resolution small-angle x-ray scattering (SAXS) data and a computational model based on chemical probing and functional data (*12*, *19*), there was no prior information on the 3D structure of BMV TLS. To assign RNA helices to regions of the cryo-EM map, we used a strategy based on RNA structural modularity. We designed BMV TLS RNAs with specific helical stems extended or truncated by several base pairs, with the rationale that local differences between the wild-type and modified RNAs would identify the extended and truncated helices within the maps. Analogous experiments have been used to validate structural models or to measure the angles between RNA helices (*20*, *21*).

A map of an RNA with stem B3 extended and stem C shortened displayed local differences compared with that of a wild type that conclusively identified B3 and C (Fig. 1D; $B3^{ext}$ + $C^{short}$). We repeated this analysis with an RNA with stem D extended and stem B2 shortened (Fig. 1D; $D^{ext}$ + $B2^{short}$). An unexpected new tertiary interaction apparently formed between the modified helices, but this did not affect the global architecture of the RNA, and local differences between the maps identified stems B2 and D. This strategy, in combination with the previously determined secondary structure (*19*) (Fig. 2A), allowed us to unambiguously identify the positions of all helices with no prior assumptions. This method of obtaining RNA helical assignments in cryo-EM maps is generally applicable and can also validate structural models from automated computational tools (*22*), as discussed below and as has been done previously (*20*).

### Cryo-EM yields a complete structural model of BMV TLS

To build a structural model consistent with the cryo-EM data and the secondary structure of BMV TLS, we first evaluated a decades-old computational 3D model (fig. S5) (*19*). The computational model correctly predicted five helical stems emanating from a central core but did not fit well into the experimental map. We built a new structure by docking individual domains into their corresponding density, performing molecular dynamics flexible fitting and real-space refinements, and correcting RNA geometry (Fig. 2B and fig. S6). The entire structure could be built and refined within the map without substantial steric clashes or breaks in the chain (table S1). Models of the extended and truncated constructs (Fig. 1D) support this structural model (fig. S7).

As a complementary method of structural modeling, we used auto-DRRAFTER, a fully automated computational tool that generates multiple unbiased models from a user-provided secondary structure and low-to-moderate–resolution cryo-EM maps (*22*). The models generated by auto-DRRAFTER agreed well with our original structure [root mean square deviation (RMSD) < 3.2 Å] (fig. S6B).

Globally, the BMV TLS structure contains three sets of coaxially stacked extended helical domains (Fig. 2C). One domain comprises helices C, B1, and B2, which form a pseudocontinuous helix that spans the structure, with the apical loops of B2 and C pointing in opposite directions (Fig. 2C, center). The second comprises helices D and A connected through a central four-way junction (4wj) (Fig. 2C, right). Helix A contains the 3′ CCA and serves as both the acceptor stem for tyrosylation and the replication initiation site (*10*). The third corresponds to the conformationally dynamic domain mentioned above (Fig. 1F) and contains helices B3 and E (hereafter referred to as B3 + E) linked to the core of the structure by a single strand of unpaired RNA (Fig. 2C, left). Thus, although atomic-level details are ambiguous at the overall map resolution, the cryo-EM structure of BMV TLS reveals functionally important features.

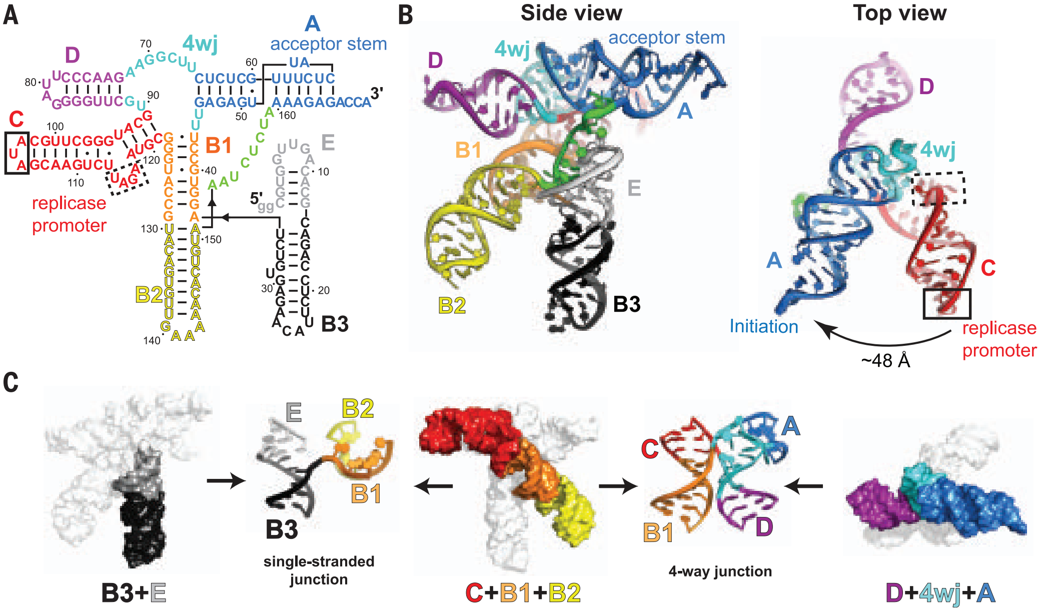


**Fig. 2. Structure of free BMV TLS RNA.** (**A**) Secondary structure of the BMV TLS RNA construct used for cryo-EM is labeled and colored by domain, with the locations of the replicase promoter and the acceptor stem labeled. The AUA apical triloop (solid box) and UAGA internal loop (dashed box) are critical for replication. (**B**) Structural model of BMV TLS from two views; colors match those in (A). Sequences critical for replication are boxed as in (A). (**C**) The helices, comprising three domains of BMV TLS formed by helical stacks, are highlighted with their corresponding colors: B3 + E (left; black and gray), C + B1 + B2 (middle; red, orange, and yellow), and D + 4wj + A (right; purple, cyan, and blue). The junctions connecting the different domains are shown between them. Colors match those in (A) and (B).

### Identification of elements that drive aminoacylation

The ability of the BMV TLS to be aminoacylated has led to the hypothesis that the rules for BMV TLS tyrosylation by TyrRS largely match those of tRNA$^{Tyr}$ (*19*, *23*). Specifically, because tRNA$^{Tyr}$ recognition by the TyrRS requires interactions with its acceptor stem and anticodon loop (with the acceptor stem being more important), analogs of these domains were expected in the BMV TLS structure (*10*, *19*, *24*, *25*). Pseudoknotted helix A was known to be the acceptor stem analog, but the identity of the putative anticodon stem analog was mysterious (*10*, *26*), with conflicting evidence for B3 or B2 (*16*, *19*, *23*, *26*). Our 3D structure shows that although B2 is oriented away from the acceptor stem, the B3 + E domain is positioned such that it is more likely to interact with TyrRS (Fig. 2B). Furthermore, a consensus sequence and secondary structural model based on conservation and covariation analysis of 512 different viral BMV-like TLS$^{Tyr}$ sequences reveals that certain nucleotide identities in B3's apical loop and the length of the B3 + E domain are highly conserved (Fig. 3, A and B). By contrast, stem B2 varies in length with no compensatory changes in the length of B1 (Fig. 3B), and there is no strong conservation in the apical loop of B2 that would suggest it acts as the anticodon (Fig. 3A). The secondary structure model derived from the covariation analysis was validated with chemical probing of three representative TLS$^{Tyr}$ variants (fig. S8). These observations strongly point to B3 as the putative anticodon stem analog.

To further investigate the role of B3, we performed in vitro tyrosylation of BMV TLS variants with terminal loops of B2, B3, or E mutated to UUCG, a loop sequence with a well-determined structure and high thermodynamic stability (*27*). In contrast to previous large sequence deletions (*16*), the UUCG mutations were not expected to affect the overall BMV TLS structure. Electrophoresis of folded BMV TLS mutants under native conditions showed unchanged migration relative to that of the wild type (fig. S9). Although mutations to B3 decreased aminoacylation of the TLS (Fig. 3C), mutations to B2 or E did not have a substantial effect, which suggests that B3 is analogous to the tRNA anticodon. Although the B3 apical loop contains a conserved UACA sequence rather than a canonical tyrosine anticodon (i.e., GUA in plants), the major identity elements of tRNA$^{Tyr}$ lay within its acceptor arm, and the anticodon is less important (*24*); this appears to also be true with BMV TLS.

### Free BMV TLS RNA does not contain a classic L-shaped tRNA mimic

To productively interact with TyrRS, it was expected that the BMV TLS acceptor and anticodon stem analogs (stems A and B3, respectively) would comprise a classic tRNA L-shaped fold (Fig. 3D, left). Unexpectedly, stem A and domain B3 + E of BMV TLS are loosely associated, with no interactions analogous to those between the T and D loops in tRNA, and the overall dimensions do not match those of tRNA (Fig. 3D). The implications were apparent when we used the crystal structure of a yeast tRNA$^{Tyr}$-TyrRS complex (*25*) to model BMV TLS bound to TyrRS (Fig. 3E). Superposition of the structure of BMV TLS on tRNA$^{Tyr}$ bound to the TyrRS homodimer, based on their acceptor stems (fig. S10), revealed substantial steric clashes between TyrRS and the B3 + E domain of BMV TLS (Fig. 3E and fig. S10B). This suggests that BMV TLS requires conformational changes to bind TyrRS, and/or binds TyrRS with a geometry that differs from that of canonical tRNA, and/or binds to a different site. The fact that B3 + E is conformationally dynamic made it plausible that interaction with TyrRS requires its rearrangement.

### BMV TLS undergoes large conformational changes to bind TyrRS

We applied cryo-EM to the BMV TLS–TyrRS complex (Fig. 4 and figs. S11 and S12). Initial studies with a 200-kV microscope (Fig. 4A) showed two copies of BMV TLS RNA (Fig. 4B, red arrows) bound to opposite sides of the TyrRS homodimer (Fig. 4B, cyan arrow), with each copy making two contacts on the enzyme. This resembles the overall configuration of the tRNA$^{Tyr}$-TyrRS complex (Fig. 4C). However, there is substantially more space between the TLS and the surface of the enzyme compared with tRNA$^{Tyr}$-TyrRS, suggesting different angles between the acceptor and anticodon stems. Consistent with our interpretation of the density, cryo-EM data of free TyrRS (fig. S13) matched the density observed in the center of the BMV TLS–TyrRS complex (Fig. 4B, cyan box).

We collected a larger dataset with a 300-kV microscope at different tilt angles to reduce the effect of preferred particle orientations (fig. S11). Although two bound BMV TLS RNAs are observed in many 2D classes (Fig. 4B and fig. S11), one RNA was consistently better defined in 3D reconstructions, which suggests compositional and/or conformational heterogeneity. Consistent with this, and as

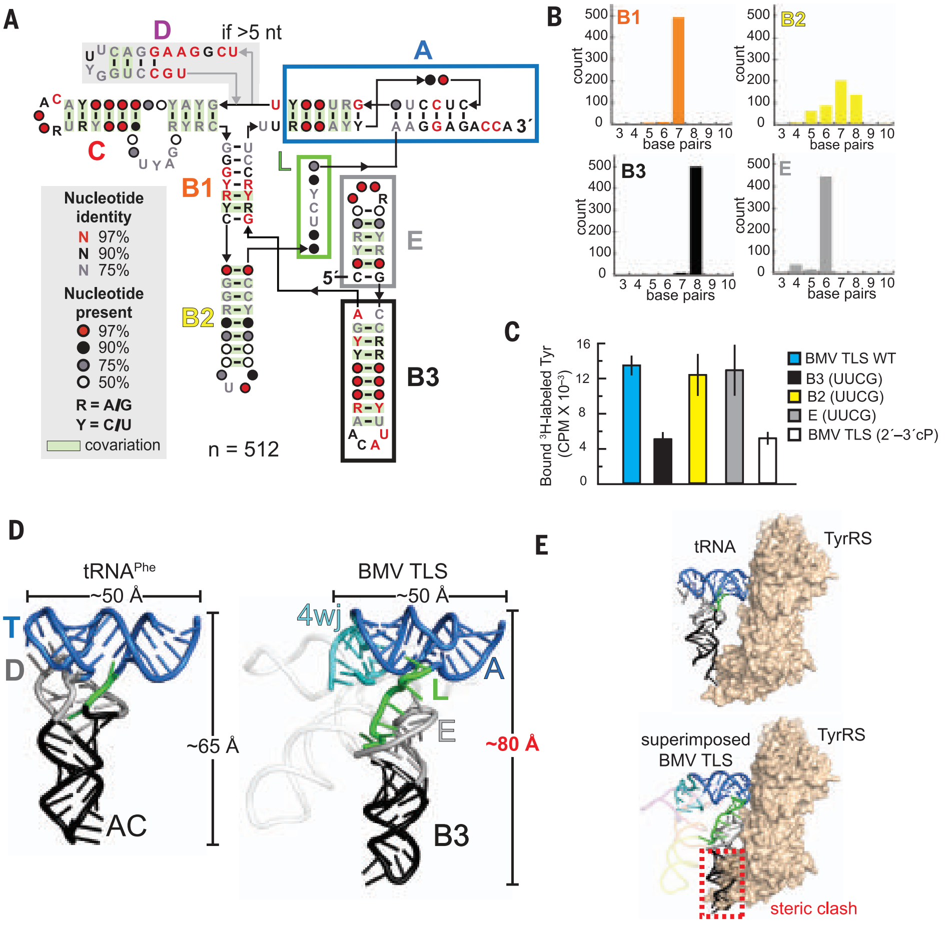


**Fig. 3. Structural features of BMV TLS important for replication or tRNA mimicry.** (**A**) Consensus covariation and sequence model of TLS$^{Tyr}$ variants related to BMV TLS. Helices are labeled according to BMV TLS. Preliminary alignment of eight TLS$^{Tyr}$ sequences from the Rfam database (*37*, *38*) followed by homology searches resulted in 512 unique sequences used in this consensus model, which is supported by chemical probing of three representative TLS$^{Tyr}$ variants (fig. S9). (**B**) Distribution of lengths of four helices according to the alignment. (**C**) Tyrosylation of mutant BMV TLS. BMV TLS (2′–3′ cP) contains a terminal 2′–3′ cyclic phosphate and thus is not an efficient substrate of TyrRS. (**D**) Comparison of unbound tRNA$^{Phe}$ (left) and tRNA mimicking portions of the BMV TLS (right). The structurally important loops T and D and anticodon (AC) loop are labeled. Analogous structural features are colored as per Fig. 2, and various molecular dimensions are shown. (**E**) (Top) Published crystal structure of yeast tRNA$^{Tyr}$ bound to yeast TyrRS (*25*). (Bottom) The cryo-EM–derived BMV TLS structure overlaid on the bound tRNA$^{Tyr}$ (fig. S11). The steric clash between B3 (anticodon stem analog) and TyrRS is boxed.

discussed below, the data revealed at least two distinct bound conformational states, bound state 1 and bound state 2 (Fig. 4, D and E, and fig. S11). Notably, a previous study had reported that although the TyrRS homodimer has two tRNA$^{Tyr}$ binding sites and crystallizes with two copies of tRNA$^{Tyr}$, only one tRNA$^{Tyr}$ is aminoacylated at a time (*28*). This apparent paradox may be reflected in the asymmetrical binding behavior of the two BMV TLS RNAs, but further studies are required.

Focusing on a single copy of BMV TLS RNA bound to TyrRS, we obtained interpretable cryo-EM 3D maps of the global conformation of the TLS bound to TyrRS (Fig. 4, D and E, top panels, and fig. S12), with resolutions of 5.5 and 6.0 Å for bound states 1 and 2, respectively (fig. S12). The maps allowed us to fit atomic models of BMV TLS RNA and TyrRS (Fig. 4, D and E, bottom panels).

The structures of the complex show that BMV TLS RNA undergoes a large conformational change to bind TyrRS (Fig. 4F). Specifically, the unbound and bound RNAs were essentially superimposable except for domain B3 + E (fig. S14). In the unbound state, B3 + E occupies a conformational ensemble, mostly positioned at roughly a right angle to the acceptor stem but not properly positioned to interact productively with the enzyme (Fig. 3E). However, in the bound state, B3 + E has rotated ~90° from its average unbound position to lay roughly parallel to the acceptor stem analog (stem A), avoiding any steric clash with the enzyme and placing the B3 apical loop on the surface of TyrRS (Fig. 4, D to F). The bound BMV TLS geometry is very different from that of tRNA$^{Tyr}$; although BMV TLS also makes two discrete contacts with TyrRS, it has little resemblance to the global L-shaped structure of tRNA$^{Tyr}$. This bound conformation was not observed in the free RNA, suggesting that it is unstable and rarely adopted without interactions with TyrRS. Thus, the BMV TLS undergoes a marked programmed conformational change to bind the synthetase, in contrast to preorganized tRNAs and TYMV TLS$^{Val}$—the latter of which nearly perfectly mimics tRNA (*29*).

## BMV TLS binds TyrRS in at least two distinct states

The BMV TLS–TyrRS complex adopts two distinct states that may relate to the process of aminoacylation. The overall conformations of the RNA and the enzyme are the same in both states (fig. S14, A and C), but their relative positions differ (Fig. 4, D and E). In bound state 1, the acceptor stem makes limited contacts with TyrRS, and the terminal CCA is well outside the aminoacylation active site (Fig. 4D). In bound state 2, the RNA and the enzyme are closer, the acceptor stem makes deeper contacts with TyrRS, and the terminal CCA is positioned in the active site. Because the position of the acceptor stem in bound state 2 more closely resembles that of tRNA$^{Tyr}$ bound to TyrRS, this state more likely reflects the bound conformation during aminoacylation. However, we cannot make conclusions about the order of events, for example, whether bound state 1 is an on-path intermediate or an alternate nonproductive state. Additionally, because the conformation of the second bound RNA was not resolved, we do not know whether all combinations of bound states 1 and 2 are present or whether there are preferred combinations. Although this behavior may be specific to BMV TLS, it is possible that similar multiple bound states exist in tRNA$^{Tyr}$-TyrRS complexes, but only a single state was observed by crystallography.

## The replicase promoter is prepositioned in proximity to the initiation site

A critical function of BMV TLS is recruiting replication machinery to the initiation site

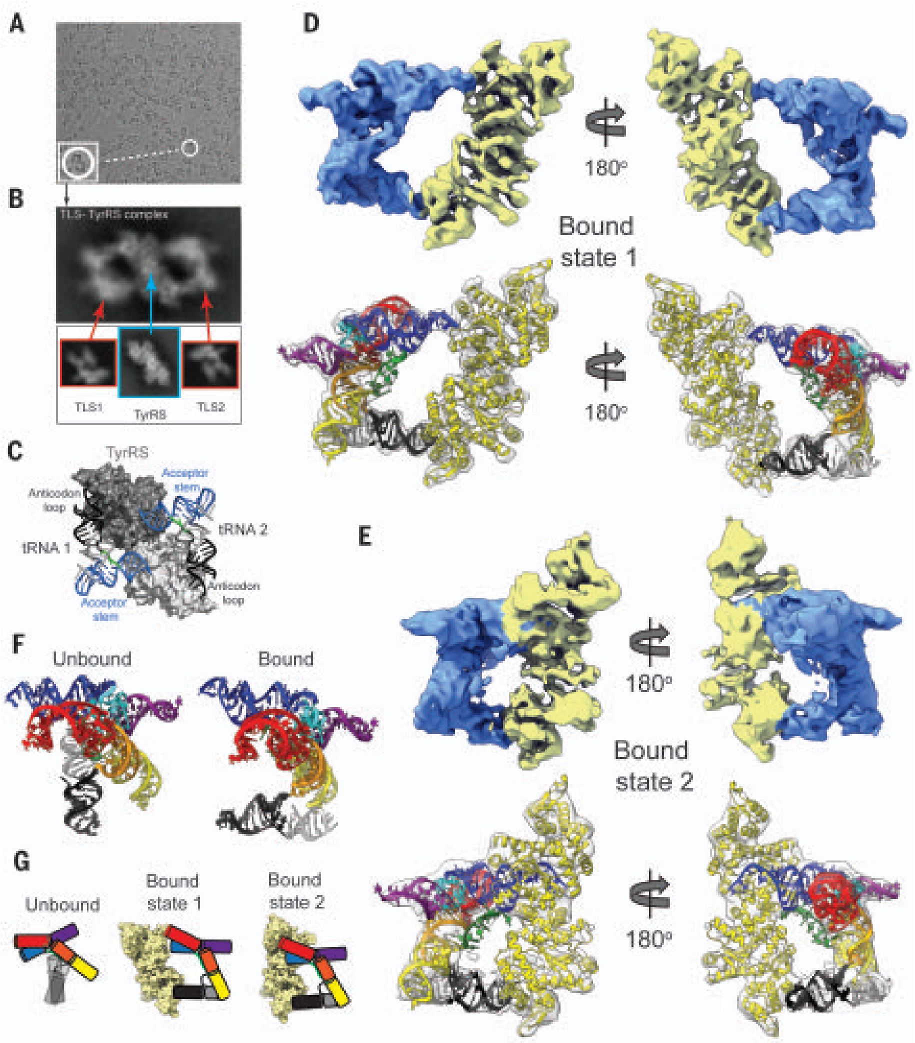


**Fig. 4. Cryo-EM of the BMV TLS–TyrRS complex.** (**A**) Representative micrograph of the BMV TLS–TyrRS complex with a 200-kV electron microscope. (**B**) Representative 2D class of the complex showing two BMV TLS RNA molecules bound to TyrRS. For comparison, densities of unbound TyrRS (cyan box; fig. S13) and unbound BMV TLS (red boxes; Fig. 1C) are shown. (**C**) Crystal structure of the yeast tRNA-TyrRS complex, with tRNA elements labeled (Protein Data Bank ID: 2dlc). (**D**) A 5.5-Å cryo-EM map of the BMV TLS–TyrRS complex in bound state 1 (top) and an atomic model fitted to the density (bottom). (**E**) A 6.0-Å cryo-EM map of the BMV TLS–TyrRS complex in bound state 2 (top) and an atomic model fitted to the density (bottom). (**F**) Comparison of the structure of BMV TLS in isolation versus when bound to TyrRS. (**G**) Comparison of BMV TLS in an unbound state and two bound states. Colors are as in Fig. 2.

(*13*, *30*, *31*). Although replication initiates at the 3′ end in stem A, the promoter is within stem C's apical AUA triloop (Fig. 2, A and B, solid box) and a UAGA 4-nt bulge (Fig. 2, A and B, dashed box) (*30*–*32*). The structure reveals that helix C is prepositioned adjacent and nearly parallel to helix A (Fig. 2B, right). The resulting distance between the AUA triloop promoter and the replication initiation site is ~48 Å. Because the replication complex consists of multiple viral and host proteins of 43 to 110 kDa (BMV-encoded replication proteins P1 and P2 are 110 and 95 kDa, respectively) (*33*), it is likely large enough to span this distance. Further, the proximity of the UAGA 4-nt bulge (Fig. 2B, dashed box) to the 4wj likely creates tertiary interactions that stabilize the position of stem C relative to helix A (Fig. 2B, right). Overall, the structure suggests that helix C and helix A are prepositioned to reduce the conformational search of the bound replicase for its substrate, but replicase binding, as with TyrRS, could also be associated with additional RNA conformational changes.

## Discussion

The BMV TLS has served as a model system for understanding RNA structure-function relationships for decades, but until now its 3D structure had remained elusive. The cryo-EM structures of the BMV TLS in unbound and TyrRS-bound states reveal a strategy for co-opting the cell's machinery using an RNA structure that contains a combination of conformationally dynamic and relatively static elements. The BMV TLS achieves aminoacylation by positioning the CCA at the 3′ end and an anticodon loop analog within an architecture that has little resemblance to that of tRNA, but which spatially arranges them to interact with the TyrRS; the decades-old term tRNA-like structure may be a misnomer for this RNA. This unexpected mode of binding invites speculation that other RNAs with secondary structures that deviate substantially from those of tRNAs may bind aaRSs in noncanonical ways but be difficult to identify on the basis of sequence or secondary structure alone.

The potential function and/or consequences of the conformational change and global fold of BMV TLS are not clear. The conformational change could serve as a signal to initiate synthetase binding or to enable the binding of other proteins to the remodeled structure. In the unbound form, the apical loop of stem E is occluded, but it is exposed in the bound form and in proximity to stem B2; any functional implications of this are unknown. The structure is a combination of preformed and dynamic features. These features might facilitate and organize interactions with the distinct machineries required for replication, recombination, and encapsidation of the viral RNAs (*30*). Other multifunctional RNAs likely utilize similar characteristics to organize different roles.

RNAs exist as conformational ensembles, and dynamics are critical for RNA function (*1*, *8*, *34*). Recent studies have explored cryo-EM as a powerful tool for rapidly solving small RNA-only structures (*22*, *35*), and our studies highlight the potential of cryo-EM for dissecting dynamic processes involving functional structured RNAs and RNA-protein complexes. Unlike crystallographic studies, in which conformational dynamics must largely be inferred or suppressed, cryo-EM offers direct detection of inherent motions. Here, we highlighted a tool that can aid in this task—the use of extensions and/or truncations of modular helical domains in combination with robust secondary structures to rapidly assign specific secondary structural elements to the electron density, provide constraints for structural modeling, and/or validate automated modeling programs. Cryo-EM, in combination with emerging computational tools, will greatly facilitate the study of diverse dynamic regulatory RNAs and RNA-protein complexes.

## REFERENCES AND NOTES

1. L. R. Ganser, M. L. Kelly, D. Herschlag, H. M. Al-Hashimi, *Nat. Rev. Mol. Cell Biol.* **20**, 474–489 (2019).
2. Z. A. Jaafar, J. S. Kieft, *Nat. Rev. Microbiol.* **17**, 110–123 (2019).

3. B. M. Akiyama, D. Eiler, J. S. Kieft, *Curr. Opin. Struct. Biol.* **36**, 40–47 (2016).
4. K. Lu *et al.*, *Science* **334**, 242–245 (2011).
5. C. P. Jones, W. A. Cantara, E. D. Olson, K. Musier-Forsyth, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3395–3400 (2014).
6. D. Yang, J. L. Leibowitz, *Virus Res.* **206**, 120–133 (2015).
7. O. Fernández-Miragall, S. López de Quinto, E. Martínez-Salas, *Virus Res.* **139**, 172–182 (2009).
8. X. Shi, S. Bonilla, D. Herschlag, P. Harbury, *Methods Enzymol.* **558**, 75–97 (2015).
9. K. Murata, M. Wolf, *Biochim. Biophys. Acta Gen. Subj.* **1862**, 324–334 (2018).
10. T. W. Dreher, *WIREs RNA* **1**, 402–414 (2010).
11. R. Giegé, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12078–12081 (1996).
12. J. A. Hammond, R. P. Rambo, M. E. Filbin, J. S. Kieft, *RNA* **15**, 294–307 (2009).
13. T. W. Dreher, *Virus Res.* **139**, 217–229 (2009).
14. R. L. Joshi, S. Joshi, F. Chapeville, A. L. Haenni, *EMBO J.* **2**, 1123–1127 (1983).
15. M. Vieweger, E. D. Holmstrom, D. J. Nesbitt, *Biophys. J.* **109**, 2625–2636 (2015).
16. P. Fechter, R. Giegé, J. Rudinger-Thirion, *J. Mol. Biol.* **309**, 387–399 (2001).
17. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, *Nat. Methods* **14**, 290–296 (2017).
18. A. Punjani, D. J. Fleet, *J. Struct. Biol.* **213**, 107702 (2021).
19. B. Felden, C. Florentz, R. Giegé, E. Westhof, *J. Mol. Biol.* **235**, 508–531 (1994).
20. N. J. Baird *et al.*, *J. Am. Chem. Soc.* **132**, 16352–16353 (2010).
21. T. M. Nakamura, Y. H. Wang, A. J. Zaug, J. D. Griffith, T. R. Cech, *EMBO J.* **14**, 4849–4859 (1995).
22. K. Kappel *et al.*, *Nat. Methods* **17**, 699–707 (2020).
23. V. Perret, C. Florentz, T. Dreher, R. Giege, *Eur. J. Biochem.* **185**, 331–339 (1989).
24. P. Fechter, J. Rudinger-Thirion, A. Théobald-Dietrich, R. Giegé, *Biochemistry* **39**, 1725–1733 (2000).
25. M. Tsunoda *et al.*, *Nucleic Acids Res.* **35**, 4289–4300 (2007).
26. T. W. Dreher, T. C. Hall, *J. Mol. Biol.* **201**, 41–55 (1988).
27. M. Molinaro, I. Tinoco Jr., *Nucleic Acids Res.* **23**, 3056–3063 (1995).
28. W. H. Ward, A. R. Fersht, *Biochemistry* **27**, 1041–1049 (1988).
29. T. M. Colussi *et al.*, *Nature* **511**, 366–369 (2014).
30. A. L. Rao, C. Cheng Kao, *Virus Res.* **206**, 46–52 (2015).
31. M. R. Chapman, C. C. Kao, *J. Mol. Biol.* **286**, 709–720 (1999).
32. C.-H. Kim, C. C. Kao, I. Tinoco Jr., *Nat. Struct. Biol.* **7**, 415–423 (2000).
33. R. Quadt, E. M. Jaspars, *Virology* **178**, 189–194 (1990).
34. L. Salmon, S. Yang, H. M. Al-Hashimi, *Annu. Rev. Phys. Chem.* **65**, 293–316 (2014).
35. K. Zhang *et al.*, *Nat. Commun.* **10**, 5511 (2019).
36. E. W. Hartwick *et al.*, *Nat. Commun.* **9**, 5074 (2018).
37. I. Kalvari *et al.*, *Curr. Protoc. Bioinformatics* **62**, e51 (2018).
38. I. Kalvari *et al.*, *Nucleic Acids Res.* **46**, D335–D342 (2018).

### ACKNOWLEDGMENTS

E. R. Camacho and P. Van Blerkom (University of Colorado Anschutz Medical Campus, Cryo-EM Facility) and T. Humphreys (Pacific Northwest Center for Cryo-EM) assisted with microscope operation. E. Hartwick helped with the aminoacylation assays. The authors thank current and former Kieft laboratory members for thoughtful discussions and technical assistance as well as A.-L. Steckelberg, B. Akiyama, Q. Vicens, and D. Constantino for critical reading of the manuscript. **Funding:** This work was supported by NIH grants R35GM118070 (J.S.K.) and F32GM139385 (S.L.B.). S.L.B. is a Howard Hughes Medical Institute Hanna Gray Fellow. M.E.S. is a Jane Coffin Childs Postdoctoral Fellow. A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. **Author contributions:** S.L.B. conceptualized the study, performed cryo-EM and aminoacylation experiments, analyzed data, performed structural modeling, and wrote the manuscript in collaboration with J.S.K. M.E.S. performed bioinformatic analysis and chemical probing and edited the manuscript. A.M. purified proteins and edited the manuscript. J.S.K. conceptualized the study, performed structural modeling, acquired funding, and wrote the manuscript in collaboration with S.L.B. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text and the supplementary materials. Cryo-EM maps are available in the Electron Microscopy Data Bank with codes EMD-24952, EMD-25023, and EMD-25041. Structural models are available in the Protein Data Bank with codes 7SAM, 7SC6, and 7SCQ.

**SUPPLEMENTARY MATERIALS**

science.org/doi/10.1126/science.abe8526
Materials and Methods
Figs. S1 to S14
Tables S1 and S2
References (*39–60*)
MDAR Reproducibility Checklist
Data S1

18 September 2020; resubmitted 27 August 2021
Accepted 4 October 2021
10.1126/science.abe8526

---

QUANTUM MECHANICS

# Quantum mechanical double slit for molecular scattering

**Haowen Zhou, William E. Perreault, Nandini Mukherjee*, Richard N. Zare***

**Interference observed in a double-slit experiment most conclusively demonstrates the wave properties of particles. We construct a quantum mechanical double-slit interferometer by rovibrationally exciting molecular deuterium ($D_2$) in a biaxial ($v = 2, j = 2$) state using Stark-induced adiabatic Raman passage, where $v$ and $j$ represent the vibrational and rotational quantum numbers, respectively. In $D_2$ ($v = 2, j = 2$) → $D_2$ ($v = 2, j' = 0$) rotational relaxation via a cold collision with ground state helium, the two coherently coupled bond axis orientations in the biaxial state act as two slits that generate two indistinguishable quantum mechanical pathways connecting initial and final states of the colliding system. The interference disappears when we decouple the two orientations of the bond axis by separately constructing the uniaxial states of $D_2$, unequivocally establishing the double-slit action of the biaxial state. This double slit opens new possibilities in the coherent control of molecular collisions.**

Studying the underlying quantum principles that drive all molecular interactions will promote our ability to manipulate the microscopic world, with potential applications ranging from molecular engineering to the realization of a quantum computer. At the heart of quantum mechanics lies the uncertainty principle, which discards the classical idea of localization, thereby introducing the wave property of matter. As a direct consequence of the wave nature of matter, interactions at the quantum scale are often controlled by interference of indistinguishable quantum mechanical pathways connecting initial and final states (*1*). The importance of this property extends all the way to biology, where interference between quantum mechanical pathways makes possible highly efficient energy transport from chromophores to the photosynthetic reaction center (*2*). We describe here the fabrication of a molecular interferometer that uses the fundamental wave character of quantum systems to interrogate and control molecular collision processes.

One of the earliest and most fundamental demonstrations of quantum mechanical interference was Young's original double-slit experiment with optical waves. Following this, a long series of experiments have been designed that use monoenergetic beams of particles ranging in mass from electrons to large molecules to show matter-wave interferences when the particles pass through two mechanical slits (*3*–*6*). These interferences only appear when the transverse coherence is maintained across both slits either by using a sufficiently large De Broglie wavelength or by restricting the size of the slit that admits the particle beam into the double-slit interferometer. More recently, double-slit interference has been realized in unimolecular and bimolecular scattering by using the two atomic centers of a diatomic molecule, which are the natural double slits of the quantum world (*7*–*10*). Although these experiments have culminated in the successful realization of the Bohr-Einstein recoiling double-slit thought experiment (*8*, *11*), the naturally occurring double slits cannot be used to actively control a molecular process. Instead, we use specially prepared quantum states to construct a double slit, which advances the coherent control of molecular scattering.

Stimulated by the recent interest in quantum science, there has been an upsurge of experimental (*12*–*18*) and theoretical (*19*–*24*) investigations on the control of molecular processes using interferences between coherently prepared quantum states. Many molecular scattering experiments have observed quantum interference that results from the material properties of the system without achieving control (*25*–*32*). On the other hand, our earlier cold scattering experiments (*16*–*18*) achieved steric control by aligning the molecules using Stark-induced adiabatic Raman passage (SARP) (*33*, *34*) by choosing different polarizations

---

Department of Chemistry, Stanford University, Stanford, CA 94305, USA.
***Corresponding author. Email: nmukherj@stanford.edu (N.M.); zare@stanford.edu (R.N.Z.)**

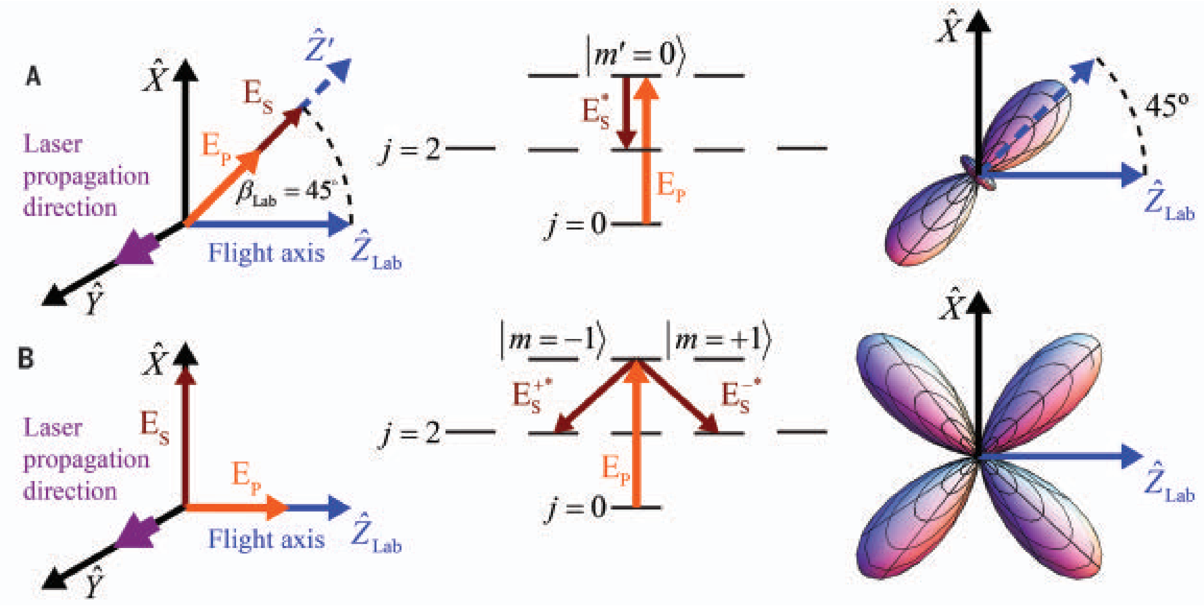


**Fig. 1. SARP preparation of uniaxial and biaxial states of $D_2$ ($v$ = 2, $j$ = 2) by combining different polarizations of the pump and Stokes laser pulses.** (**A**) SARP preparation of $D_2$ ($v$ = 2, $j$ = 2) with bond axis aligned at $\beta_{Lab} = \pi/4$ relative to the lab-fixed axis $Z_{Lab}$ oriented along the molecular beam axis, which is also the flight axis of the detected ions. This alignment was achieved using parallel pump and Stokes linear polarizations along $Z'$ oriented at 45° relative to $Z_{Lab}$. The quantum number $m'$ refers to the projection of the $j$ = 2 angular momentum vector along the quantization axis $Z'$. (**B**) SARP preparation of the biaxial state with cross-polarized pump (along $Z_{Lab}$) and Stokes (along $X$) laser pulses. The $m$ quantum number refers to the projection of the $j$ = 2 angular momentum vector along the quantization axis $Z_{Lab}$. In both schematics, $E_P$ refers to the absorption of a pump photon, and $E_S^*$ refers to the emission of a Stokes photon in a lambda-type transition. On the right, we display the three-dimensional molecular axis distributions for the prepared states of $D_2$.

of the pump and Stokes lasers (*35*). Each of these alignments can be expressed as a superposition of the magnetic sublevels within a rovibrational energy state, and so the measured scattering distribution results from interference between different sublevels. Others have performed similar experiments, including, for example, Nichols *et al.* (*12*), who studied orientational effects in inelastic collision by aligning the NO molecules using a dc field. Alignment in a strong electric field produced a coherent mixture of the even and odd Λ- doublet states, which created interference effects in the output of the scattering process. However, each of these superpositions did not form a double slit because, with a suitable choice of the basis set, it could be converted to a single quantum state (uniaxial system). We achieve a double-slit interference in a two-body collision by creating coherent coupling between two orthogonal states of one of the collision partners.

Specifically, we studied the rotational relaxation of aligned $D_2$ ($v$ = 2, $j$ = 2) → $D_2$ ($v$ = 2, $j'$ = 0) at a single collision level, mediated by a ground state He atom within a cold supersonically expanded mixed molecular beam, where $v$ and $j$ represent the vibrational and rotational quantum numbers, respectively. SARP is able to transfer nearly the complete ground state population of $D_2$ to the rovibrational energy eigenstate ($v$ = 2, $j$ = 2) (*36*) with its bond axis prepared in a biaxial orientation in the laboratory frame (*35*). The biaxial state coherently couples two orientations of the $D_2$ bond axis, meaning that the molecule simultaneously exists in two possible orientations, and so no coordinate transformation can convert it to a pure state. As such, the density matrix for this state cannot be diagonalized by any unitary transformation. Thus, defining this nonclassical state requires the phase coherence information represented by the nonzero off-diagonal density matrix in addition to population information. The two orientations generate two quantum mechanical pathways in the rotationally inelastic collision that interfere to produce a strong modulation in the scattering intensity distribution. As such, scattering from the coupled orientations of the biaxial state mimics the way that a photon or particle simultaneously experiences the two mechanical slits in a traditional double-slit scattering experiment. Our earlier theoretical study (*37*) showed that the biaxial state is capable of producing an interference pattern in the scattering intensity in the plane perpendicular to the collision velocity. In this study, we experimentally demonstrate the double-slit action and show that the interference is evident in the scattering distribution projected parallel to the collision velocity. To establish unequivocally the double-slit action of the biaxial state, we show that the strong modulation in the scattering intensity distribution disappears when the contributions from the two orientations are considered independently, just as the interference pattern disappears in a traditional double slit by blocking one of the slits. Thus, our experiment not only illustrates the ability of quantum interference to control a molecular collision but also shows that the biaxial state is a purely quantum mechanical object that has a one-to-one correspondence with a double slit.

### Experimental demonstration of a molecular double slit

To study cold collisions with SARP-prepared uniaxial and biaxial states, we expanded a mixture of $D_2$ and He (ratio 1:10) in a single supersonic beam using an Even-Lavie pulsed valve (*38*). The supersonically expanded and collimated beam (divergence ≤12 mrad) not only dramatically reduced the collision temperature (~1 K) (*39*) but also defined a direction with respect to which the $D_2$ bond axis is aligned using SARP. Using resonance-enhanced multiphoton ionization (REMPI), we determined that ~40% of the $D_2$ ($v$ = 0) population in the mixed supersonic beam was in the $j$ = 0 state, with the remainder distributed among the $j$ = 1, 2, and 3 states. To prepare the uniaxial and biaxial states of $D_2$, we intersected the $D_2$-He mixed molecular beam transversely with a sequence of pump and Stokes laser pulses partially overlapping in time. A comprehensive description of SARP can be found elsewhere (*34*). The SARP preparation process is able to transfer nearly 100% of the $D_2$ ($v$ = 0, $j$ = 0) population to the target $D_2$ ($v$ = 2, $j$ = 2) state

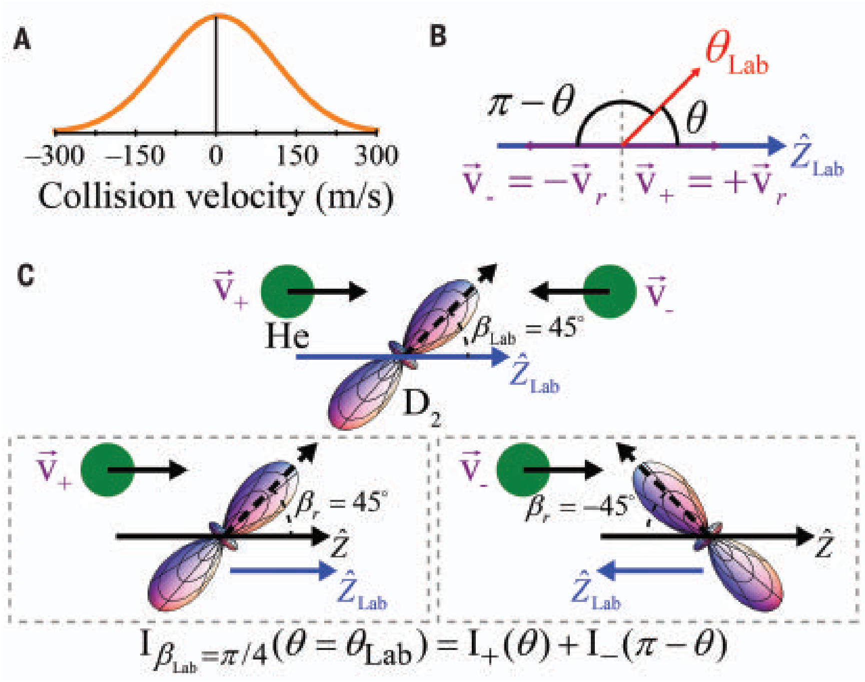


**Fig. 2. Effect of the symmetric collision velocity distribution on the scattering measurement.** (**A**) Experimentally determined collision velocity distribution for the $D_2$-He scattering in the mixed molecular beams. (**B**) Schematic representation of the conversion of center-of-mass-frame scattering angle to lab-frame scattering angle for the positive ($\vec{v}_+$) and negative ($\vec{v}_-$) velocities relative to $Z_{\text{Lab}}$. (**C**) Schematic demonstration that for the $\beta_{Lab} = \pi/4$ preparation, the $\vec{v}_+$ velocity group sees a center-of-mass orientation $\beta_r = \pi/4$, whereas the $\vec{v}_-$ velocity group sees a center-of-mass orientation $\beta_r = -\pi/4$.

(*36*). After SARP preparation, the $D_2$ molecules collided with He atoms and relaxed to the ($v$ = 2, $j'$ = 0) state, where they were state-selectively ionized using REMPI.

The orientation of the molecular bond axis was controlled using the polarization directions of the pump and Stokes laser pulses and is defined relative to a lab-fixed frame with coordinate axes $X$, $Y$, and $Z_{\text{Lab}}$. Figure 1 describes the experimental arrangement and gives a schematic of SARP excitation to prepare the uniaxial (Fig. 1A) and biaxial (Fig. 1B) states. Figure 1A shows the preparation of $D_2$ ($v$ = 2, $j$ = 2) with its bond axis oriented at $\beta_{\text{Lab}} = \pi/4$ relative to $Z_{\text{Lab}}$ using pump and Stokes fields with polarization along $\beta_{\text{Lab}} = \pi/4$. In a similar way, $D_2$ ($v$ = 2, $j$ = 2) can be prepared with its bond axis aligned at $\beta_{\text{Lab}} = -\pi/4$ relative to $Z_{\text{Lab}}$ using pump and Stokes laser polarizations along $\beta_{\text{Lab}} = -\pi/4$ (not shown). We call the wave functions of the two uniaxial states with the bond axis aligned at $\beta_{\text{Lab}} = \pi/4$ and $\beta_{\text{Lab}} = -\pi/4$ as $\psi_+$ and $\psi_-$, respectively. Figure 1B shows the preparation of a biaxial state using SARP with a cross-polarized pump and Stokes laser fields. This state is called XSARP, and its wave function is denoted by $\psi_X$. We note that biaxial and multiaxial atomic states have been prepared for use in scattering experiments previously (*40*, *41*), but this work did not examine interference effects as we will do here.

Using the angular momentum representation in terms of $m$, the projection quantum number of the angular momentum vector $\boldsymbol{j}$ on the quantization axis $Z_{\text{Lab}}$, the wave functions of the two uniaxial states $\psi_+$ and $\psi_-$ are expressed as

$$\psi_\pm = 0.25|m=0\rangle \pm 0.612[|m=-1\rangle - |m=+1\rangle] + 0.306[|m=+2\rangle + |m=-2\rangle] \quad (1)$$

In the same representation, the XSARP can be expressed as

$$\psi_X = (|m=+1\rangle - |m=-1\rangle)/\sqrt{2} \quad (2)$$

The three states, $\psi_\pm$ and $\psi_X$ are defined with respect to the same quantization axis $Z_{\text{Lab}}$. Using Eqs. 1 and 2, $\psi_X$ can be written as a superposition of the two uniaxial states $\psi_\pm$:

$$\psi_X = 0.577\left[|\psi_+\rangle - |\psi_-\rangle\right] \quad (3)$$

The biaxial state $\psi_X$ thus coherently ties the two axes of the uniaxial states $\psi_+$ and $\psi_-$. The data we present here show that this phase coherence causes a strong modulation in the intensity distribution of the scattered $D_2$ ($v$ = 2, $j'$ = 0) in the same fashion as a double slit causes interference fringes.

The double-slit interference effect in the scattering of the biaxial state can be readily seen in the expression for the center-of-mass-frame angular distribution for the XSARP state. For the biaxial XSARP state in Eq. 3, each rotationally inelastic collision is mediated by the $D_2$ bond axis simultaneously occupying two alignments at $\beta_{\text{Lab}} = \pi/4$ and $\beta_{\text{Lab}} = -\pi/4$. The scattering intensity distribution as a function of the polar angle $\theta$ is obtained by integrating over the azimuthal angle $\varphi$. The scattering intensity for $\psi_X$ is thus

$$I_X(\theta) = \sin\theta\left[\int_0^{2\pi} |f_X(\theta,\varphi)|^2 d\varphi\right] \quad (4)$$

where $f_X$ represents the scattering amplitude for $D_2$ prepared in the initial state $\psi_X$. Using the expression of $\psi_X$ in Eq. 3, we can write

$$f_X(\theta,\varphi) = 0.577[f_+(\theta,\varphi) - f_-(\theta,\varphi)] \quad (5)$$

In Eq. 5, $f_+$ and $f_-$ give the center of mass scattering amplitudes for the initial quantum states prepared in $\psi_+$ and $\psi_-$, respectively. Using Eq. 5, the scattering intensity for the biaxial state can be expressed as follows:

$$I_X(\theta) = [I_+(\theta) - I_-(\theta) - I_{\text{Int}}(\theta)] \quad (6)$$

where $I_\pm(\theta)$ represents the center-of-mass-frame scattering angular distribution for the initial state prepared in $\psi_\pm$ and is given by

$$I_\pm(\theta) = \sin\theta\int_0^{2\pi} |f_\pm(\theta,\varphi)|^2 d\varphi \quad (7)$$

$I_\pm(\theta)$ gives the contribution of individual axis orientations, or slits, and is entirely determined by the absolute square of the scattering amplitude $f_\pm$. $I_{\text{Int}}(\theta)$ gives rise to the double-slit interference

$$I_{\text{Int}}(\theta) = \sin\theta\int_0^{2\pi} \left[f_+(\theta,\varphi)f_-^*(\theta,\varphi) + f_+^*(\theta,\varphi)f_-(\theta,\varphi)\right] d\varphi \quad (8)$$

and is determined from the relative phase of the scattering amplitudes $f_+$ and $f_-$. The interference term arises from coherence between the two bond axis orientations, which is equivalent to the phase correlation between the two slits in the double-slit experiment when the De Broglie wavelength becomes sufficiently large to couple both the slits. This interference manifests itself as strong modulation in the scattering intensity, as we show below.

However, the lab-frame angular distribution that we measure and show below is not the same as the center-of-mass-frame distribution. This is because collisions with both positive ($\vec{v}_+ = \vec{v}_r$) and negative ($\vec{v}_- = -\vec{v}_r$) relative velocities with respect to the $+Z_{\text{Lab}}$ axis are

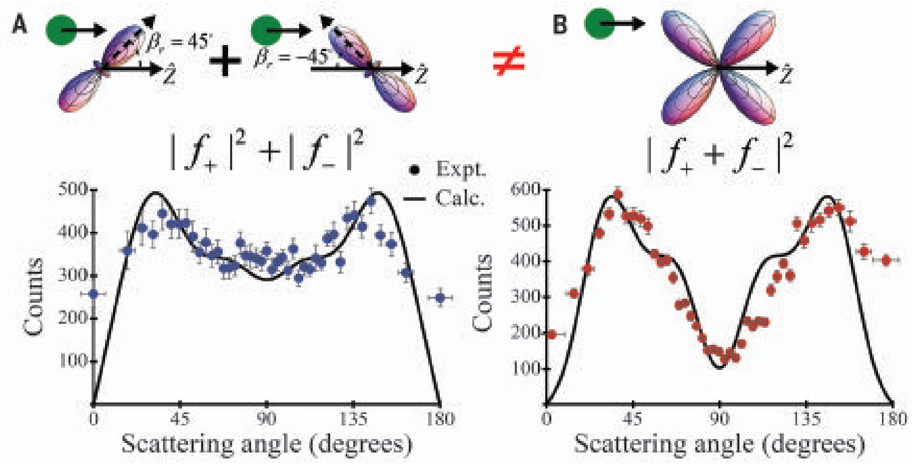


**Fig. 3. Demonstration of double slit by preparing biaxial and uniaxial states for the bond axis of $D_2$ ($v$ = 2, $j$ = 2).** For clarity, the top panels only show the collision for the positive collision velocity group. (**A**) Experimental scattering distribution (blue dots) for the uniaxial bond axis orientation at $\beta_{\rm Lab} = \pi/4$. As discussed in the text, this is identical to the distribution measured with the $\beta_{\rm Lab} = -\pi/4$ orientation. (**B**) Experimental scattering distribution (red dots) for the XSARP biaxial bond axis orientation. The black solid lines in (A) and (B) show the calculated angular distribution using the scattering matrix. The vertical error bars represent one standard deviation in the number of counts at a given scattering angle, and the horizontal error bars give the experimental uncertainty in the determination of the scattering angle.

present in the symmetric collision velocity distribution shown in Fig. 2A. As a result, the experimentally measured scattering intensity at a lab angle $\theta_{\rm Lab} = \theta$ defined relative to $+Z_{\rm Lab}$, is the sum of the contributions due to scattering at an angle θ from the positive ($\vec{v}_+$) collision velocity and scattering at angle π – θ from the negative ($\vec{v}_-$) collision velocity. As shown in Fig. 2C, for the negative velocity group, the center-of-mass-frame $Z$ axis is flipped with respect to the lab frame. As a result, the bond axis orientation in the center-of-mass frame for the negative collision velocity group is flipped such that $\beta_r = -\beta_{\rm Lab}$. Including the contributions of both collision velocity groups, the lab-frame scattering intensities for the uniaxial states $\psi_+$ and $\psi_-$ are given by

$$I_{\beta_{\rm Lab}=\pm\pi/4}(\theta = \theta_{\rm Lab}) = I_{\pm}(\theta) + I_{\mp}(\pi - \theta) \quad (9)$$

Using Eq. 9, we arrive at the following expression for the lab-frame scattering angular distribution for the biaxial state:

$$I_X(\theta_{\rm Lab} = \theta) = \{I_+(\theta) + I_-(\pi - \theta) + I_-(\theta) + I_+(\pi - \theta) - [I_{\rm Int}(\theta) + I_{\rm Int}(\pi - \theta)]\} = \{I_{\beta_{\rm Lab}=\pi/4}(\theta) + I_{\beta_{\rm Lab}=-\pi/4}(\theta) - [I_{\rm Int}(\theta) + I_{\rm Int}(\pi - \theta)]\} \quad (10)$$

We have used Eqs. 9 and 10 to analyze the experimental angular distribution for the uniaxial and biaxial states shown in Fig. 3. In Eq. 10, the contributions of $I_{\beta_{\rm Lab}=\pm\pi/4}$ represent the transmission of one slit at a time. This contribution is experimentally isolated because we were able to prepare each of the uniaxial states separately, and so we can show explicitly the double-slit interference of the biaxial state that arises from the cross terms $f_+(\theta, \varphi) f_-^*(\theta, \varphi)$ in the expression of $I_{\rm Int}$ (Eq. 8). For the homonuclear symmetry of $D_2$, we can show analytically that the φ integrated scattering intensities for the two uniaxial states are identical, which we have also verified experimentally. The supplementary materials contain a description of these experiments, as well as a more in-depth derivation of Eqs. 9 and 10.

Figure 3 compares the experimental scattering angular distribution for the uniaxial and biaxial bond axis orientation of $D_2$ ($v$ = 2, $j$ = 2). Our experimental method for measuring scattering angular distributions has been described in detail in an earlier publication (*16*). Figure 3A shows the experimental scattering data obtained using the uniaxial alignment at $\beta_{\rm Lab} = \pi/4$ As we mentioned above and show in detail in the supplementary materials, the scattering angular distribution for $\beta_{\rm Lab} = -\pi/4$ was identical. Figure 3B shows the measured scattering intensity for the biaxial state. Clearly, the double-slit interference that is present in the scattering of the biaxial state produced a strong modulation in the scattering intensity near 90°. Using the scattering matrix for the rotationally inelastic scattering $D_2$ ($v$ = 2, $j$ = 2) → $D_2$ ($v$ = 2, $j'$ = 0), which was determined in our previous experiments, we calculated the scattering intensities using Eqs. 9 and 10 (see supplementary materials for details). The calculated scattering intensity is superimposed on the experimental data shown in Fig. 3 for the uniaxial and biaxial alignments. The strong correspondence between the calculations and our experimental measurements not only demonstrates the double-slit action of the biaxial state but also validates the scattering matrix elements that we determined in our earlier work (*15*).

## Conclusion

We have created a quantum mechanical double slit by preparing the rovibrationally excited $D_2$ ($v$ = 2, $j$ = 2) molecule in a biaxial state. The coherently coupled bond axis orientations in the biaxial state lead to two indistinguishable pathways connecting the input state and final state of the $\Delta j$ = 2 rotationally inelastic $D_2$-He collision. Therefore, they act as the two slits of a double-slit interferometer. This double-slit interference manifested as a strong modulation in the measured angular distribution of the rotationally relaxed $D_2$ ($v$ = 2, $j'$ = 0). To prove that the observed interference arose from the double-slit action of the biaxial state, we measured the effect of one slit at a time by separately constructing each of the uniaxial states contained within the biaxial state superposition. Measuring the contributions of the individual slits separately is equivalent to decoupling the axes in the biaxial state. When we add the scattering angular distributions measured for each slit, the result is notably different from that of the biaxial state. This difference that clearly arises from the quantum mechanical interference of the two orientations, which act as the two slits, unequivocally establishes the double-slit action of the biaxial state.

As discussed in the introduction, many previous atomic and molecular scattering experiments have used the quantum mechanical interference of a double slit to interpret their results. Of these, only a few were able to experimentally isolate the double-slit contribution from that of decoupled single slits and thereby demonstrate that interference arises exclusively when the two slits are coupled as we have done here. One elegant example (*11*) examined the interferences present in the Auger spectrum of a core-excited diatomic molecule that arise as a result of two indistinguishable pathways for electron emission when the excited electron is delocalized over the two atomic centers. They showed that the interference disappears when the system is prepared in the dissociative state because it becomes possible to identify which slit (atom) the electron is emitted from. As impressive as these experiments are, there is unfortunately not a straightforward way to use these double slits as a tool to control further scattering processes.

In the present work, we have demonstrated the use of quantum states as slits as opposed to individual atoms. This gives us the ability to use our experimentally engineered molecular quantum interferometer to control or measure the fundamental properties of a molecular process. For example, using our molecular double slit, the relative phases of the scattering matrix elements can be experimentally measured (*42*), which allows a direct test of theoretical calculations on important, fundamental collision processes. In this example, we have demonstrated fabrication of a double slit by coherently coupling the two uniaxial states with a fixed phase relationship. A higher level of control can be obtained by introducing a variable phase relation between the coherently coupled states that can be further manipulated using various optical and electro-optical techniques (*35*). Our ability to coherently control the molecular double slit is not limited to inelastic collisions but is more general, opening completely new possibilities in the control of bimolecular reactions.

## REFERENCES AND NOTES

1. R. P. Feynman, R. B. Leighton, M. L. Sands, *Feynman Lectures on Physics, Vol. 3* (Addison Wesley Publishing Co., 1963).
2. G. S. Engel *et al.*, *Nature* **446**, 782–786 (2007).
3. A. H. Tavabi *et al.*, *Sci. Rep.* **9**, 10458 (2019).
4. S. Frabboni, G. C. Gazzadi, G. Pozzi, *Am. J. Phys.* **75**, 1053–1055 (2007).
5. M. Arndt *et al.*, *Nature* **401**, 680–682 (1999).
6. O. Carnal, J. Mlynek, *Phys. Rev. Lett.* **66**, 2689–2692 (1991).
7. D. Misra *et al.*, *Phys. Rev. Lett.* **102**, 153201 (2009).
8. L. P. H. Schmidt *et al.*, *Phys. Rev. Lett.* **111**, 103201 (2013).
9. D. Akoury *et al.*, *Science* **318**, 949–952 (2007).
10. S. E. Canton *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7302–7306 (2011).
11. X. J. Liu *et al.*, *Nat. Photonics* **9**, 120–125 (2015).
12. B. Nichols *et al.*, *Chem. Sci.* **6**, 2202–2210 (2015).
13. M. H. G. de Miranda *et al.*, *Nat. Phys.* **7**, 502–507 (2011).
14. S. D. S. Gordon *et al.*, *Nat. Chem.* **10**, 1190–1195 (2018).
15. H. Zhou, W. E. Perreault, N. Mukherjee, R. N. Zare, *J. Chem. Phys.* **154**, 104309 (2021).
16. W. E. Perreault, N. Mukherjee, R. N. Zare, *Science* **358**, 356–359 (2017).
17. W. E. Perreault, N. Mukherjee, R. N. Zare, *J. Chem. Phys.* **150**, 174301 (2019).
18. W. E. Perreault, N. Mukherjee, R. N. Zare, *Nat. Chem.* **10**, 561–567 (2018).
19. M. Shapiro, P. Brumer, *Phys. Rev. Lett.* **77**, 2574–2576 (1996).
20. A. Devolder, T. Tscherbul, P. Brumer, *Phys. Rev. A* **102**, 031303 (2020).
21. P. Brumer, A. Abrashkevich, M. Shapiro, *Faraday Discuss.* **113**, 291–302 (1999).
22. J. L. Krause, M. Shapiro, P. Brumer, *J. Chem. Phys.* **92**, 1126–1131 (1990).
23. C. A. Arango, M. Shapiro, P. Brumer, *J. Chem. Phys.* **125**, 094315 (2006).
24. P. Král, I. Thanopulos, M. Shapiro, *Rev. Mod. Phys.* **79**, 53–77 (2007).
25. G. Sha, J. He, B. Jiang, C. Zhang, *J. Chem. Phys.* **102**, 2772–2779 (1995).
26. J. Onvlee *et al.*, *Nat. Chem.* **9**, 226–233 (2017).
27. X. L. Chen *et al.*, *Chem. Phys. Lett.* **318**, 107–112 (2000).
28. P. G. Jambrina *et al.*, *Nat. Chem.* **7**, 661–667 (2015).
29. P. G. Jambrina, J. Aldegunde, F. J. Aoiz, M. Sneha, R. N. Zare, *Chem. Sci.* **7**, 642–649 (2016).
30. M. Sneha *et al.*, *J. Chem. Phys.* **145**, 024308 (2016).
31. Y. Xie *et al.*, *Science* **368**, 767–771 (2020).
32. B. K. Kendrick *et al.*, *Phys. Chem. Chem. Phys.* **23**, 5096–5112 (2021).
33. N. Mukherjee, R. N. Zare, *J. Chem. Phys.* **135**, 024201 (2011).
34. N. Mukherjee, W. E. Perreault, R. N. Zare, in *Frontiers and Advances in Molecular Spectroscopy*, J. Laane, Ed. (Elsevier, 2018), pp. 1–46.
35. N. Mukherjee, W. Dong, R. N. Zare, *J. Chem. Phys.* **140**, 074201 (2014).
36. W. E. Perreault, N. Mukherjee, R. N. Zare, *J. Chem. Phys.* **150**, 234201 (2019).
37. W. E. Perreault, H. Zhou, N. Mukherjee, R. N. Zare, *Front. Phys. (Lausanne)* **9**, 671997 (2021).
38. U. Even, *EPJ Tech. Instrum.* **2**, 17 (2015).
39. W. E. Perreault, N. Mukherjee, R. N. Zare, *Chem. Phys.* **514**, 150–153 (2018).
40. R. L. Robinson, L. J. Kovalenko, S. R. Leone, *Phys. Rev. Lett.* **64**, 388–391 (1990).
41. J. P. J. Driessen, C. J. Smith, S. R. Leone, *J. Phys. Chem.* **95**, 8163–8169 (1991).
42. Y. Gao *et al.*, *Phys. Rev. A* **97**, 020701 (2018).
43. H. Zhou, W. E. Perreault, N. Mukherjee, R. N. Zare, Quantum mechanical double slit for molecular scattering. Dataset. Dryad (2021); https://doi.org/10.5061/dryad.jh9w0vtcb.

## ACKNOWLEDGMENTS

**Funding:** This work was supported by the US Army Research Office through the MURI program under grant no. W911NF-19-1-0283 and the National Science Foundation under grant no. PHY- 2110256. **Author contributions:** Experimental data were taken by H.Z., W.E.P., and N.M., and calculations were performed by N.M. All authors participated in discussion and writing of the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in this paper are present in the paper or supplementary materials. All data presented in this paper are deposited at Dryad (*43*).

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4143
Materials and Methods
Supplementary Text
Figs. S1 and S2
Table S1
References (*44*, *45*)

12 July 2021; accepted 23 September 2021
10.1126/science.abl4143

---

REPORTS

JUPITER

# The depth of Jupiter's Great Red Spot constrained by Juno gravity overflights

**Marzia Parisi[1]*, Yohai Kaspi[2], Eli Galanti[2], Daniele Durante[3], Scott J. Bolton[4], Steven M. Levin[1], Dustin R. Buccino[1], Leigh N. Fletcher[5], William M. Folkner[1], Tristan Guillot[6], Ravit Helled[7], Luciano Iess[3], Cheng Li[8]†, Kamal Oudrhiri[1], Michael H. Wong[8,9]**

**Jupiter's Great Red Spot (GRS) is the largest atmospheric vortex in the Solar System and has been observed for at least two centuries. It has been unclear how deep the vortex extends beneath its visible cloud tops. We examined the gravity signature of the GRS using data from 12 encounters of the Juno spacecraft with the planet, including two direct overflights of the vortex. Localized density anomalies due to the presence of the GRS caused a shift in the spacecraft line-of-sight velocity. Using two different approaches to infer the GRS depth, which yielded consistent results, we conclude that the GRS is contained within the upper 500 kilometers of Jupiter's atmosphere.**

Jupiter's Great Red Spot (GRS) is a large and enduring anticyclonic storm in Jupiter's atmosphere. Both its size (currently more than 16,000 km east-west) and centuries-old longevity are unlike other vortices in the Solar System and must be driven by the underlying dynamics of the storm. It is unknown whether the vortex is deep-rooted in the surrounding zonal flows (east-west bands), or whether a shallow system can maintain such a long-living storm (*1*, *2*). Spacecraft gravity measurements are sensitive to density perturbations in the deeper layers of the atmosphere, below the visible cloud level, so close encounters with the vortex provide an opportunity to study the GRS gravity signature and infer the depth of its winds.

The Juno spacecraft flew twice over the GRS, in February and July 2019, at altitudes below 20,000 km. These encounters (perijoves) were dedicated to measurements of the local gravitational field. Earlier observations of the GRS had occurred during another perijove in July 2017, during which the spacecraft attitude was optimized for Microwave Radiometer (MWR) investigations that precluded gravity measurements. The microwave observations indicated that the GRS extends down to at least 100-bar atmospheric pressure, well below the cloud level (estimated at 0.7 bar), which corresponds to a depth of ~240 km (*3*). MWR observations are complementary to gravity measurements because they provide a minimum value for the vortex's vertical extension (*3*). Previous Juno gravity measurements have

---

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA. [2]Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot 76100, Israel. [3]Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, 00184 Rome Italy. [4]Southwest Research Institute, San Antonio, TX 78238, USA. [5]School of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. [6]Université Côte d'Azur, Observatoire de la Côte d'Azur, Laboratoire Lagrange, Centre National de la Recherche Scientifique, 06304 Nice, France. [7]Center for Theoretical Astrophysics and Cosmology, Institute for Computational Science, University of Zurich, 8057 Zurich, Switzerland. [8]Department of Astronomy, University of California, Berkeley, CA 94720, USA. [9]Carl Sagan Center for Research, SETI Institute, Mountain View, CA 94043, USA.
***Corresponding author. Email: marzia.parisi@jpl.nasa.gov**
†Present address: Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109, USA.

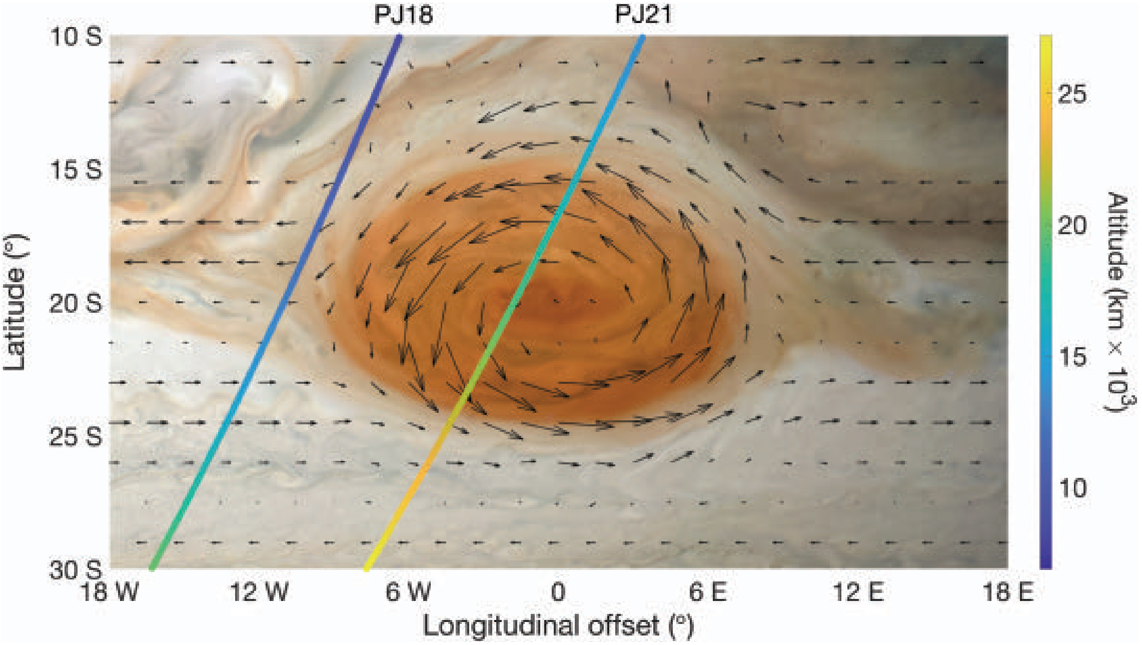


**Fig. 1. Geometry of the Juno GRS gravity observations.** The GRS velocity field (black arrows) at the time of PJ21 (*20*, *32*) with Juno tracks during PJ18 and PJ21 (gradient lines) is superimposed on JunoCam imaging of the GRS during PJ21 (*33*). The spacecraft altitude during the GRS closest approach (latitude 20°S) was 13,000 km during PJ18 and 19,000 km during PJ21, with eastward longitudinal offsets of 11° and 2°, respectively. [Background image credit: NASA/JPL-Caltech/SwRI/MSSS/Kevin M. Gill © CC BY, released under the Creative Commons Attribution 3.0 Unported (CC BY 3.0) license https://creativecommons.org/licenses/by/3.0/legalcode]

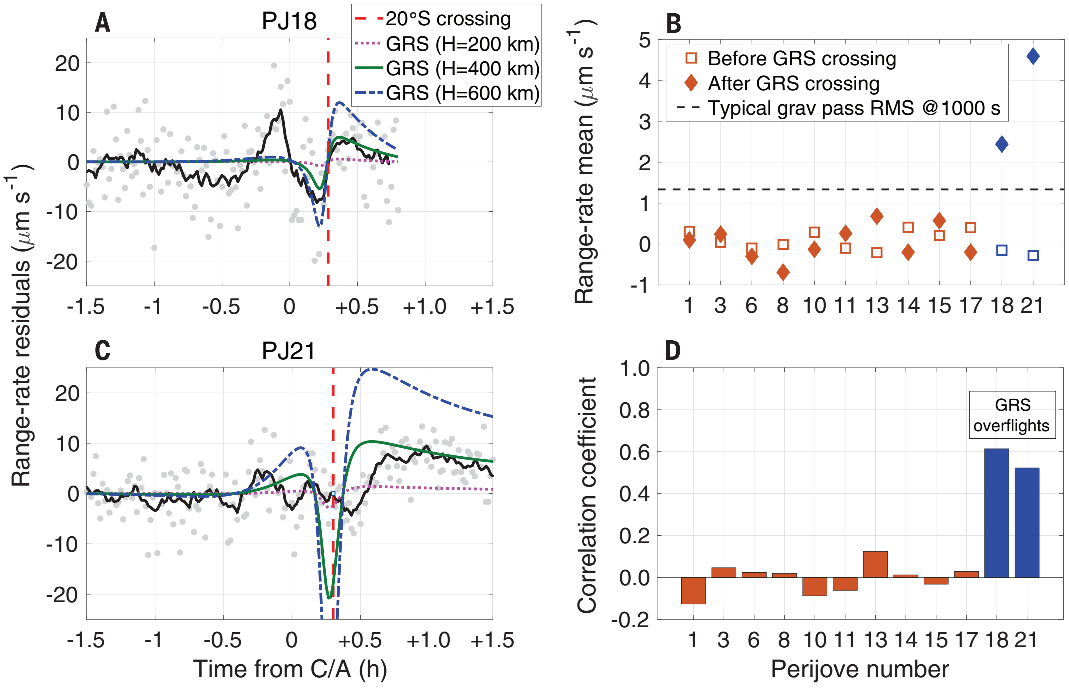


**Fig. 2. Nonzonal, nonstatic effects on the range-rate residuals from PJ18 and PJ21.** (**A** and **C**) PJ18 (A) and PJ21 (C) range-rate residuals (gray dots) with their 10-min moving average (black solid line), for the period ±1.5 hours around closest approach (C/A) to Jupiter. The red dashed vertical line marks the GRS crossing (20°S) by the spacecraft track. The other lines represent the expected gravity signal from the GRS for different depths, as indicated. (**B**) Range-rate mean values ($\Delta v$), before (squares) and after (diamonds) the crossing of latitude 20°S, for all perijoves with gravity measurements. (**D**) Correlation coefficients between the predicted GRS gravity signal (for $H$ = 300 km) and the moving average of the residuals. In (C) and (D), non-GRS perijoves are orange; GRS overflights are blue.

shown that Jupiter's gravity field is hemispherically asymmetric (*4*), which has been explained by the powerful zonal winds in which the GRS is embedded (*5*). By relating the observed cloud-level winds to the gravity measurements, the depth of the jet streams has been estimated at 3000 km (*6*). This estimate assumed thermal wind balance, which is expected to dominate vorticity balance on large, rapidly rotating planets (*6*). However, it remains unclear how far down the GRS winds extend between the surrounding jet streams.

Before Juno arrived at Jupiter, Parisi *et al.* proposed to use gravity data to determine the depth of the GRS (*7*) by determining the planet's longitudinally dependent gravity field. Later on, an alternative proposal suggested the use of a Slepian basis set (*8*), whose functions are defined as linear combinations of spherical harmonics chosen to maximize the information in a selected domain. The Slepian approach has previously been used to describe localized features on Earth (*9*, *10*) and to analyze incomplete data collected over a limited area (*11*).

Simulations predicted that the gravity signal of the GRS could potentially be described with a single Slepian coefficient, and that the gravity data would be sensitive to the depth of the GRS if it was at least a few hundred kilometers (*8*). Similar predictions were reached using a complementary approach that models the GRS as a vertical dipole of mass concentrations (mascons), which estimates the mass of the anomalies directly (*12*).

Juno's orbit around Jupiter is quasi-polar and highly eccentric, with the spacecraft descending to altitudes of about 3500 km over Jupiter's cloud tops every 53 days, sweeping the planet from its north pole to its south pole within a couple of hours (*13*). The oblateness of the planet causes a northward shift of the perijove latitude by ~1° per orbit, from an initial latitude of 4°N. Therefore, the spacecraft observes the GRS (20°S) from an increasing distance as the mission proceeds. During gravity passes, two-way Doppler measurements of the spacecraft range rate are performed using coherent radio links with NASA's Deep Space Network in X-band (~8 GHz) and Ka-band (~32 GHz) (*14*), which are sensitive to perturbations in the gravity field of Jupiter. We analyzed data from 12 Juno perijoves (PJs) labeled with progressive numbers 1, 3, 6, 8, 10, 11, 13, 14, 15, 17, 18, and 21. We initialized Jupiter's global gravity field and tidal response using Juno gravity data up to PJ17 (*15*), to which we added the GRS overflights data from PJ18 and PJ21. Figure 1 shows the Juno tracks over Jupiter's reference surface during the PJ18 and PJ21 passes.

We combined the Juno gravity measurements using a least-squares multiarc filter implemented in the orbit determination software MONTE (*16*), which was also used for trajectory integration and data analysis. The range-rate integration time was 60 s and the typical noise RMS (root mean square) was between 5 and 10 $\mu$m s$^{-1}$ (one-way). The filter requires the accurate reconstruction of the Juno trajectory, including every acceleration acting on the spacecraft, both gravitational and nongravitational (*4*, *15*). We searched for the GRS gravity signature in the Juno overflight data by processing the range-rate residuals while solving for the zonal gravity field of Jupiter. Previous analyses (*15*) have discussed the difficulty of fitting the Juno range-rate data from multiple perijoves without allowing for longitudinal or temporal variations of the gravity field.

The range-rate residuals for PJ18 and PJ21, after removing the effect of the zonal gravity field, show non-negligible signals over time scales of ±1 hour around closest approach to Jupiter (Fig. 2, A and C). These deviations from zonal symmetry can be removed by the addition of short-lasting, constant accelerations (*15*) but could also be partly attributed to the GRS. The closest approaches to Jupiter occurred at latitudes around 20°N; Juno crossed the GRS latitude 20°S about 20 min later. We analyzed the remaining non–axially symmetric, nonstatic signatures on the range-rate residuals after removing the effect of the zonal gravity field of Jupiter. Our objective was to isolate the longitudinal or temporal effects, which include the signal from the GRS.

The measured PJ18 residuals (Fig. 2A) show deviations from the mean value ~10 min before perijove (latitude ~30°N). We do not expect these fluctuations to be related to the GRS, as the spacecraft was then ~100,000 km away. Figure 2A also shows the expected gravity signal from the GRS, assuming that the cloud-level winds decay at a depth $H$. The correlation between the predicted GRS signature and the smoothed residuals is shown in Fig. 2D.

PJ21 data (Fig. 2C) also show correlation with the GRS signal, with a constant shift in the line-of-sight velocity of ~5 $\mu$m s$^{-1}$, lasting until at least 1.5 hours after perijove. This behavior is typical of passing one or more mass concentrations, such as those associated to the GRS. The long duration of the constant shift favors decoupling of the constant accelerations applied around perijove from the GRS signal. We interpret the temporal offset between the negative peaks around the GRS crossing (measured versus predicted) as due to unrelated nonzonal effects. We investigated the role of the accelerations in removing the excess longitudinal and/or temporal signatures and aliasing with the GRS parameters (*17*). After compensating for the estimated accelerations (fig. S7, C and D), we found that (i) most of the remaining range-rate signal can be explained by concentrated masses at the GRS location, and (ii) unexplained nonzonal effects (*15*) can cause the offset between the PJ21 range-rate residuals and the predicted GRS signal (Fig. 2C).

The average velocity shift ($\Delta v$) in the residuals before and after the crossing of latitude 20°S, for all perijoves, is shown in Fig. 2B. These averages were calculated over time periods of 0.5 to 1.5 hours. Passes over the GRS show shifts of a few $\mu$m s$^{-1}$, above the typical noise RMS for integration times exceeding 1000 s (*17*). Conversely, the velocity profiles for perijoves away from the GRS do not show deviations from the mean. The magnitude of the offset on the range-rate data contains direct information on the GRS depth (*17*), which is reflected in the output parameters of the least-squares solution used to infer the vertical extent of the winds.

Figure 2D shows the correlation coefficients between the predicted GRS gravity signal for $H$ = 300 km and the moving average of the range-rate residuals, for all 12 perijoves (*17*). The non-GRS passes are characterized by very low correlations of less than 10%, whereas the correlations for PJ18 and PJ21 are 50 to 60%. Although the signal-to-noise ratio is low, perijoves 18 and 21 differ from the other passes, indicating that the GRS depth can be constrained using the Juno gravity data.

The shape of the GRS has evolved over the past several decades, with its longitudinal dimension shrinking and giving the vortex a more circular shape (*18*, *19*). We measured horizontal wind speeds in the GRS using optical data acquired about 25 days before PJ21 (*17*, *20*), taken as part of an annual Hubble Space Telescope observing program (*21*). Mean velocities measured in the high-speed ring are 106 m s$^{-1}$, with a standard deviation of 11 m s$^{-1}$. Remote observations provide information about the surface dynamics of the zonal winds and the vortex, but little is known about the dynamics below the cloud level. The surrounding jets extend very deep (*6*), which confines

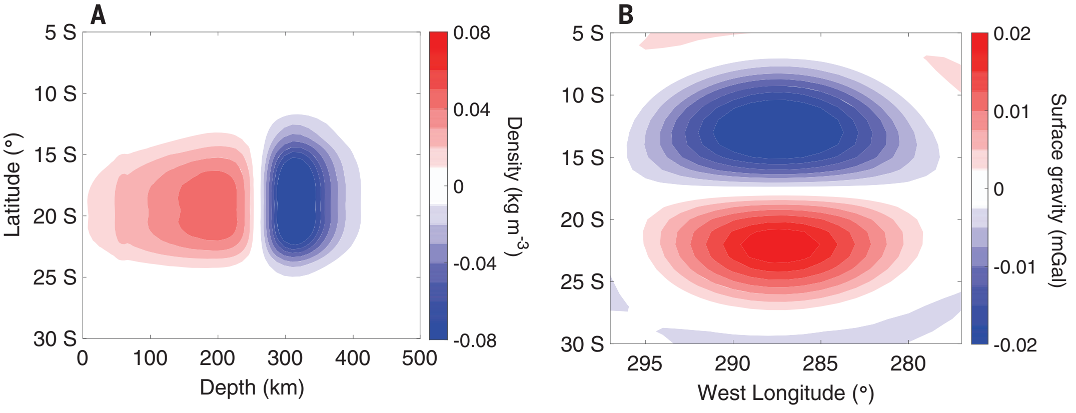


**Fig. 3. The predicted signal for a 300-km-deep GRS, assuming thermal wind balance.** (**A**) Density anomalies as a function of depth and latitude, for a transversal section taken at the longitude passing through the GRS center at the time of the velocity measurements (*17*, *20*, *32*). (**B**) Surface gravity anomalies as a function of longitude and latitude.

the latitudinal extent and direction of the GRS circulation. We therefore assume that the vortex preserves its shape until it decays below the depth $H$. For simplicity we consider a hyperbolic tangent decay function, assumed to decay rather sharply within 100 km around $H$ (*22*, *23*), but our results are robust to other choices of the decay shape (*17*). Because the planet is rapidly rotating and the dynamics are geostrophic (to leading order), thermal wind balance can be used to calculate the density anomalies balancing the vortex velocity (*24*). The applicability of this approach to modeling Jupiter's atmosphere has been discussed elsewhere (*25*, *26*). Unlike Earth's atmospheric vortices (*27*), the local centripetal force can be neglected.

The predicted density anomalies associated with the GRS for a sample depth $H$ = 300 km are shown in Fig. 3A. The density profile resembles a dipole, with a positive mascon in the upper levels [with gravitational parameter $GM_A$ (where $G$ is the gravitational constant and $M_A$ is the positive mass anomaly)] and a negative mascon at depth (with parameter $GM_B$, where $M_B$ is the negative mass anomaly). The sum of the two masses is zero, to first order (*17*). The relationship between the mass and the separation between the positive and negative anomaly is injective (fig. S3); therefore, the depth ($H$) can be inferred by using MONTE to estimate the mass (*12*). In the orbit determination software, we model the vortex as a pair of flat disk mascons, whose masses are constrained to be equal and opposite in sign. Deeper winds entail a larger mass involved in the circulation of the GRS and an increasing vertical distance between the mascons. A sign inversion between the upper and deeper levels was also observed in MWR measurements during PJ7 (*3*).

An alternative approach to searching for the GRS gravity signature is to use Slepian functions to characterize the wind-induced, concentrated surface gravity anomalies (*8*). Figure 3B shows that the GRS predicted gravity disturbances form a north-south dipole. The Slepian functions are defined within the bounded domain delimited by an ellipsoid centered at the GRS location spanning 20° in latitude and 30° in longitude. Our analysis of the Slepian functions (*17*) shows that a single function, labeled $\mathbf{g}_2$, can describe the gravity perturbations generated by the GRS, to leading order (*8*). The magnitude of the corresponding Slepian coefficient $\alpha_2$ increases with the depth of the GRS (Fig. 4B); therefore, the depth ($H$) can be inferred from a measurement of $\alpha_2$.

To account for unmodeled accelerations (Fig. 2, A and C), we introduced small constant accelerations, each of 10-min duration, for ±1 hour around perijove (*15*, *28*). In addition to standard estimated parameters, such as Jupiter's zonal gravity field and pole position (*17*), we also added one coefficient specific to the determination of the depth of the GRS: the gravitational parameter of the positive disk mascon $GM_A$ for the mascon approach, or the Slepian coefficient $\alpha_2$ for the Slepian approach. Both parameters were allowed to vary without constraint.

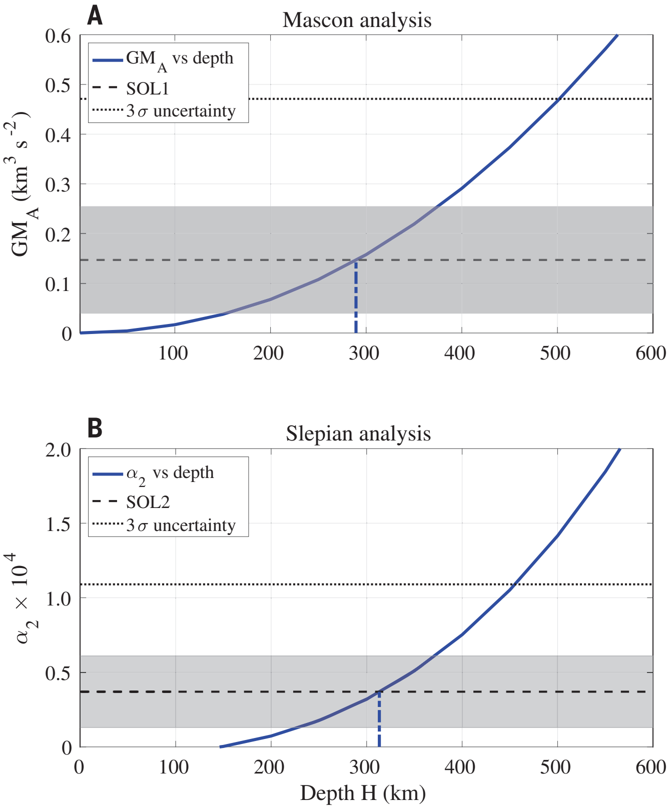


**Fig. 4. Results of the GRS depth estimation.** (**A**) Results of the mascon approach (best-fitting value labeled SOL1), shown with the predicted relationship between $H$ and $GM_A$ assuming thermal wind balance (blue). (**B**) Results of the Slepian approach (SOL2) shown with the relationship between $H$ and $\alpha_2$ for the predicted gravity anomalies (blue). In both panels, the estimated value is shown with its 1σ uncertainty (gray shaded area) and the 3σ upper limit (dotted horizontal line).

The results of the mascon analysis are shown in Fig. 4A. The dipole structure is implemented in MONTE, and the vertical separation affects both the estimated central value of $GM_A$ and its formal uncertainty (*12*). We measure $GM_A = 1.47\ (\pm 1.08) \times 10^{-1}\ \mathrm{km}^3\ \mathrm{s}^{-2}$ (all uncertainties are 1σ). The derived GRS depth is then obtained from the predicted relationship between $H$ and $GM_A$ from thermal wind balance, finding $H = 290^{+85}_{-140}$ km. For the Slepian approach (Fig. 4B), the estimated $\alpha_2$ coefficient is $3.7\ (\pm 2.4) \times 10^{-5}$, which implies that the winds extend down to $H = 310^{+60}_{-90}$ km. The two different methods give consistent solutions for the depth of the GRS and are compatible with the depth inferred from the MWR observations (*3*). We tested the stability of both solutions against different models of the unknown nonzonal, nonstatic effects, which indicated that the estimated depth is robust (*17*).

Both methods assume that the GRS is in thermal wind balance, and each has strengths and limitations, which are complementary to one another (*12*, *17*). They differ in the way the predicted density profile is used: either to model mass concentrations or to model the gravitational potential at the spacecraft altitude. They provide consistent results and indicate (Fig. 4) a 3σ upper limit on the GRS depth of 500 km (750 bar). This upper limit is compatible with laboratory analog experiments and numerical simulations (*29*, *30*). With 1σ uncertainty, we find that the GRS is ~300 ± 100 km deep. However, the minimum depth of the GRS is not well constrained by our analysis. This was expected, because the GRS gravity signal is only 5% as strong as the background zonal wind signal (*4*, *15*). Our use of random accelerations de-weights the Doppler data and increases the formal uncertainties. However, Juno MWR observations provide a minimum value for the vortex's vertical extension of ~240 km (*3*), which complements our gravity measurements. Although it is possible that the GRS winds still increase below the cloud level (*31*), before they begin decaying deeper down, any increase must be less than 50% of the cloud-level velocity, otherwise the inferred depth would be shallower than indicated by MWR. We therefore conclude that the depth of the GRS is between 200 and 500 km.

Our results suggest that the GRS is much shallower than the surrounding zonal jets, which have depths of ~3000 km (*6*). The GRS is nonetheless deeply rooted, extending far below the cloud level at 0.7 bar and well beyond the water condensation level (~80 km beneath the cloud level) (*31*). Although Jupiter does not have a solid surface, the GRS is still shallow in terms of the aspect ratio between the vertical and horizontal scales (~1/200 or 0.5%). This is even shallower than Earth's cyclones and anticyclones, which have typical ratios of 1 to 4%, limited by the depth of Earth's troposphere. The driving mechanisms for Earth's vortices are very different, with roles played by atmospheric instability processes and the solid surface, the latter not being present on Jupiter. It remains unclear why the GRS has a depth of only a few hundred kilometers while the surrounding jets, which power the GRS, extend much deeper. However, a shallow GRS is consistent with its change in size over the past several decades (*19*).

## REFERENCES AND NOTES

1. T. E. Dowling, A. P. Ingersoll, *J. Atmos. Sci.* **46**, 3256–3278 (1989).
2. J. Yano, G. R. Flierl, *Ann. Geophys.* **12**, 1–18 (1994).
3. S. J. Bolton *et al*., *Science* 10.1126/science.abf1015 (2021).
4. L. Iess *et al*., *Nature* **555**, 220–222 (2018).
5. Y. Kaspi, *Geophys. Res. Lett.* **40**, 676–680 (2013).
6. Y. Kaspi *et al*., *Nature* **555**, 223–226 (2018).
7. M. Parisi, E. Galanti, S. Finocchiaro, L. Iess, Y. Kaspi, *Icarus* **267**, 232–242 (2016).
8. E. Galanti *et al*., *Astrophys. J. Lett.* **874**, L24 (2019).
9. C. Harig, F. J. Simons, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19934–19937 (2012).
10. F. J. Simons, F. A. Dahlen, M. A. Wieczorek, *SIAM Rev.* **48**, 504–536 (2006).
11. F. J. Simons, F. A. Dahlen, *Geophys. J. Int.* **166**, 1039–1061 (2006).
12. M. Parisi *et al*., *Planet. Space Sci.* **181**, 104781 (2020).
13. S. J. Bolton *et al*., *Science* **356**, 821–825 (2017).
14. S. W. Asmar *et al*., *Space Sci. Rev.* **213**, 205–218 (2017).
15. D. Durante *et al*., *Geophys. Res. Lett.* **47**, e2019GL086572 (2020).
16. S. Evans *et al*., *CEAS Space J.* **10**, 79–86 (2018).
17. See supplementary materials.
18. A. A. Simon *et al*., *Astron. J.* **155**, 151 (2018).
19. A. Sánchez-Lavega *et al*., *Astron. J.* **156**, 162 (2018).
20. M. H. Wong *et al*., *Geophys. Res. Lett.* **48**, e2021GL093982 (2021).
21. A. A. Simon, M. H. Wong, G. Orton, *Astrophys. J.* **812**, 55 (2015).
22. E. Galanti *et al*., *Geophys. Res. Lett.* **46**, 616–624 (2019).
23. Y. Kaspi *et al*., *Space Sci. Rev.* **216**, 84 (2020).
24. Y. Kaspi, W. B. Hubbard, A. P. Showman, G. R. Flierl, *Geophys. Res. Lett.* **37**, L01204 (2010).
25. K. Zhang, D. Kong, G. Schubert, *Astrophys. J.* **806**, 270 (2015).
26. E. Galanti, Y. Kaspi, E. Tziperman, *J. Fluid Mech.* **810**, 175–195 (2017).
27. J. R. Holton, *An Introduction to Dynamic Meteorology* (Elsevier/Academic Press, 2004).
28. L. Iess *et al*., *Science* **364**, eaat2965 (2019).
29. D. Lemasquerier, G. Facchini, B. Favier, M. Le Bars, *Nat. Phys.* **16**, 695–700 (2020).
30. R. K. Yadav, M. Heimpel, J. Bloxham, *Sci. Adv.* **6**, 46 (2020).
31. L. N. Fletcher *et al*., *Icarus* **208**, 306–328 (2010).
32. M. Wong, "Jupiter Great Red Spot Velocity Fields from HST/WFC3" (Mikulski Archive for Space Telescopes); https://doi.org/10.17909/t9-jfs3-p240.
33. K. M. Gill, "Great Red Spot, Perijove 21"; www.missionjuno.swri.edu/junocam/processing?id=7258.

## ACKNOWLEDGMENTS

We thank D. J. Stevenson, A. P. Ingersoll, and the Juno Interiors and Atmospheric Working Groups for useful discussions; citizen scientist K. M. Gill, who processed the background image in Fig. 1 as part of Juno's JunoCam outreach program; and N. Gavriel for graphical assistance with Fig. 1. **Funding:** This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology (M.P., S.M.L., D.R.B., W.M.F., K.O.) under contract 80NM0018D0004 with NASA; at the Weizmann Institute of Science (WIS) in Israel (Y.K. and E.G.) with support from the Israeli Space Agency and the Helen Kimmel Center for Planetary Science at WIS; at Sapienza University of Rome (D.D. and L.I.) under contract 2017-40-H.0 with the Italian Space Agency (ASI); at the Southwest Research Institute (S.J.B.) under contract to NASA; at the University of Leicester (L.N.F.) with support by a European Research Council Consolidator Grant (under the European Union's Horizon 2020 research and innovation program, grant agreement 723890); at the Observatoire de la Côte d'Azur (T.G.) under the sponsorship of the Centre National d'Etudes Spatiales; with support by the 51 Pegasi b Postdoc Fellowship sponsored by the Heising-Simons Foundation (C.L.); at UC Berkeley and the SETI Institute (M.H.W.) with support from NASA through Participating Scientist grant 80NSSC19K1265 and through the Space Telescope Science Institute (for programs GO-13937 and GO-15502), which is operated by the Association of Universities for Research in Astronomy Inc. under NASA contract NAS 5-26555. **Author contributions:** M.P. led the data analysis and interpretation of the gravity data. M.P., Y.K., and E.G. wrote the manuscript. Y.K. proposed the concept for this study and assisted in the interpretation of the results. E.G. performed the thermal wind analysis. E.G. and D.D. performed the Slepian approach calculations. D.R.B., W.M.F., and K.O. assisted with the data collection, processing, and analysis. S.J.B., S.M.L., L.N.F., T.G., R.H., L.I., and C.L. assisted with the interpretation of the results. M.H.W. acquired the Hubble observations and determined the GRS velocity profile. **Competing interests:** M.H.W. is also affiliated with the University of Michigan. **Data and materials availability:** The Juno gravity data are available in the NASA Planetary Data System at https://atmos.nmsu.edu/PDS/data/jnogrv_1001//DATA/RSR/; we used the files from 2016240 to 2019202. The geometry of the Juno orbit from 2016 to 2019, including SPK trajectory files and CK spacecraft attitude files, is available at https://naif.jpl.nasa.gov/pub/naif/JUNO/kernels/. The raw Hubble Space Telescope imaging data are available from the Mikulski Archive for Space Telescopes (MAST) at https://archive.stsci.edu/hst/search.php under proposal ID 15502; we used datasets iduy09mmq, iduy09mvq, iduy10n4q, and iduy15ooq. The velocity field derived from those images is available in HDF5 format at MAST (*32*). Distribution of the MONTE navigation code is restricted by the Export Administration Regulations of the US Department of Commerce. Eligible readers may request a copy of MONTE, under a license that does not permit redistribution, at https://montepy.jpl.nasa.gov/. The derived range-rate residuals for PJ1 to PJ21 and partial derivatives for sample GRS parameters are available at https://doi.org/10.5281/zenodo.4649459. Executable Matlab code for the thermal wind model, for both the mascon and Slepian approaches, is available at https://github.com/egalanti/TW_GRS_Slepians.

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf1396
Materials and Methods
Figs. S1 to S9
Tables S1 to S3
References (*34–36*)
Movie S1

6 October 2020; accepted 28 April 2021
Published online 28 October 2021
10.1126/science.abf1396

---

JUPITER

# Microwave observations reveal the deep extent and structure of Jupiter's atmospheric vortices

**S. J. Bolton[1]*, S. M. Levin[2], T. Guillot[3], C. Li[4], Y. Kaspi[5], G. Orton[2], M. H. Wong[6], F. Oyafuso[2], M. Allison[7,8], J. Arballo[2], S. Atreya[4], H. N. Becker[2], J. Bloxham[9], S. T. Brown[2], L. N. Fletcher[10], E. Galanti[5], S. Gulkis[2], M. Janssen[2], A. Ingersoll[11], J. L. Lunine[12], S. Misra[2], P. Steffes[13], D. Stevenson[11], J. H. Waite[1], R. K. Yadav[9], Z. Zhang[11]**

**Jupiter's atmosphere has a system of zones and belts punctuated by small and large vortices, the largest being the Great Red Spot. How these features change with depth is unknown, with theories of their structure ranging from shallow meteorological features to surface expressions of deep-seated convection. We present observations of atmospheric vortices using the Juno spacecraft's Microwave Radiometer. We found vortex roots that extend deeper than the altitude at which water is expected to condense, and we identified density inversion layers. Our results constrain the three-dimensional structure of Jupiter's vortices and their extension below the clouds.**

Observations of Jupiter's atmosphere have shown that it is characterized by a long-lasting (more than a century) system of zones and belts disrupted by storms and vortices, including the Great Red Spot (GRS). Jupiter's weather layer—where storms, vortices, and convective clouds are observed—was expected to be constrained to depths at which sunlight can penetrate and/or water condenses. Contrary to this expectation, data from the Galileo atmospheric entry probe showed that water was not well mixed, even well below its expected condensation level (*1*). Results from the Juno spacecraft showed that both ammonia and water vary across most of the planet at much greater depths than their expected saturation levels (*2*–*4*) and that the gravitational signatures of the atmospheric zonal flows extend to depths approaching 3000 km (*5*, *6*). The depth of atmospheric vortices is expected to depend on how Jupiter's weather is constrained by volatile condensation, and the relative importance of moist convection, baroclinic instability, and deep convection in vortex creation and stability.

The Microwave Radiometer (MWR) instrument (*2*, *7*) on Juno is a set of radiometers designed to measure Jupiter's emitted flux (or equivalently the brightness temperature) at a range of depths, from the top of the

---

[1]Southwest Research Institute, San Antonio, TX, USA. [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. [3]Université Côte d'Azur, Observatoire de la Côte d'Azur, Centre National de la Recherche Scientifique, Laboratoire Lagrange, Nice, France. [4]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA. [5]Weizmann Institute of Science, Rehovot, 76100, Israel. [6]Carl Sagan Center for Research, SETI Institute, Mountain View, CA, USA. [7]Goddard Institute for Space Studies, New York, NY, USA. [8]Department of Astronomy, Columbia University, New York, NY 10027, USA. [9]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA. [10]School of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. [11]Department of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA. [12]Department of Astronomy, Cornell University, Ithaca, NY, USA. [13]Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.
***Corresponding author. Email: sbolton@swri.edu**

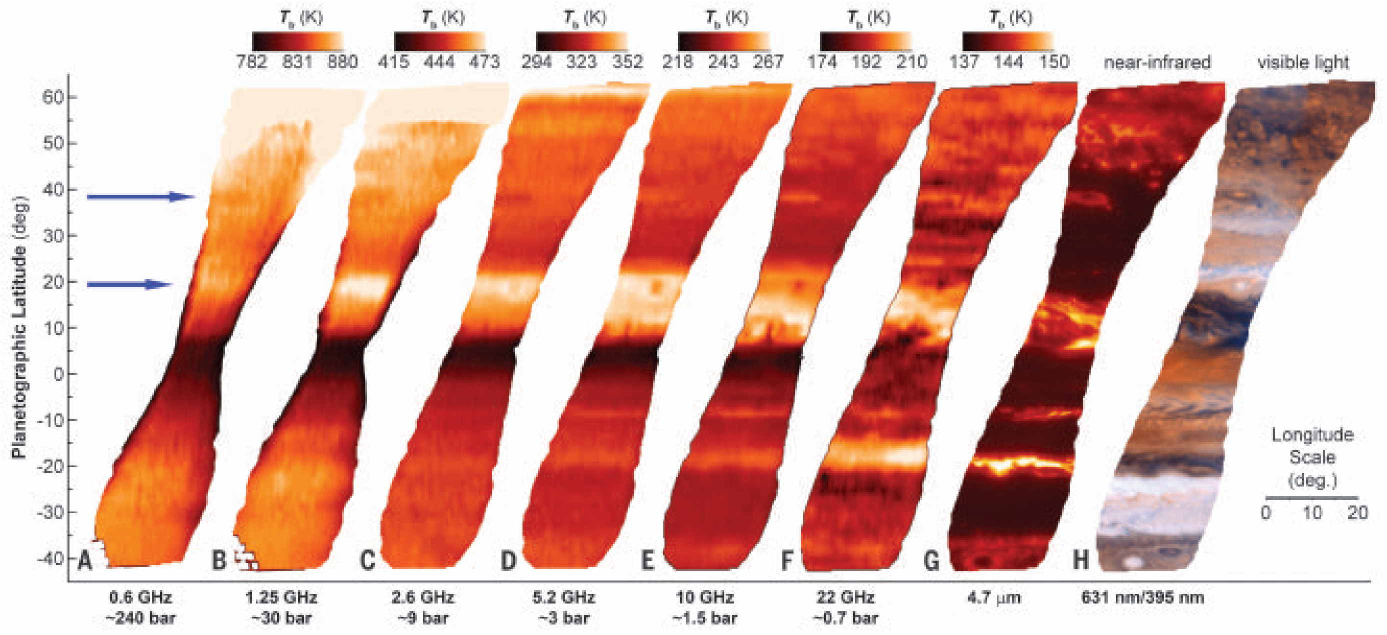


**Fig. 1. Maps of Jupiter's atmosphere at radio, IR, and optical wavelengths.** (**A** to **F**) Maps of the cross-track MWR observations, taken at frequencies labeled below each map. Also labeled are estimated pressures at which the physical temperature is equal to the average brightness temperature (*2–4*, *7*). (**G**) Near-IR and (**H**) visible light images are from the Gemini North telescope and Hubble Space Telescope, respectively (*13*). All maps were taken on 6 April 2019 and cover identical longitudes and planetographic latitudes. Blue arrows indicate the latitudes of the 19°N and 38°N vortices. The MWR resolution varies with position (fig. S1).

atmosphere to more than 600 km beneath the visible cloud tops. The instrument observes at six frequencies between 0.6 and 22 GHz (wavelengths 50 to 1.3 cm), each sampling a different depth determined by how atmospheric transparency varies with frequency (*8*). Jupiter emits more energy than it receives from the Sun, and the atmospheric temperature increases with depth (pressure). We investigated the vertical structure of two vortices observed on 6 April 2019, during Juno's 19th orbit of Jupiter (designated PJ19, for perijove 19). During PJ19, Juno was placed into a mapping mode, in which the spacecraft spin direction was turned to sample along longitude (instead of the usual latitude). This observing mode was designed to provide radio image maps as a function of frequency. Each frequency map corresponds to a different depth determined by atmospheric opacity, which is primarily due to water and ammonia. Juno's orbit passes beneath Jupiter's radiation belts, so MWR observes atmospheric thermal emission without substantial contamination by synchrotron emission from the planet's magnetosphere. Because the PJ19 scans sample across longitudes, the measurements were taken at different emission angles and thus require a limb-darkening correction. To estimate the nadir brightness temperature ($T_b$), we corrected the data for limb darkening using a model based on measurements from previous Juno orbits (*9*). The observed brightness temperature at each frequency is the result of integration of the emission from all atmospheric layers, weighted by the contribution function for each layer at the corresponding frequency (fig. S6). The resulting radio maps allow us to compare the vertical structures of Jupiter's cyclones and anticyclones, including the GRS, which was observed by Juno on 11 July 2017. A cyclonic vortex has winds blowing counterclockwise in the northern hemisphere and clockwise in the southern hemisphere; an anticyclonic vortex has winds blowing in the opposite direction.

The multifrequency MWR maps from PJ19 are shown in Fig. 1, compared with infrared (IR) 5-µm and visible-wavelength maps obtained from Earth on the same day (*8*). The higher MWR frequencies correspond to thin slices at shallow depths, and lower frequencies correspond to progressively thicker and deeper slices (fig. S6) (*2–4*, *7*, *8*). The shallowest MWR slice (22 GHz) is emitted from a slice centered where the atmospheric pressure is 0.7 bar (about 10 km above the 1-bar level, which is conventionally regarded as the surface of Jupiter); this roughly corresponds to the condensation level of ammonia ice, known as the ammonia cloud-deck. At frequencies of 5.2 to 10 GHz, MWR maps pierce through the water condensation level at ~5 to 8 bar (about 50 to 75 km below the 1-bar level). At 600 MHz, MWR's lowest frequency, the slices become thicker and deeper, reaching more than 100 bar (up to 600 km beneath the 1-bar level). Remote observations from Earth at 8 to 22 GHz previously detected a vortex known as Oval BA and the GRS near ~5 bar, although smaller vortices were not seen at these frequencies, leading to the conclusion that the small vortices are confined to the weather layer in the upper 2 to 3 bar of the atmosphere (*10*, *11*). Our observations extend to lower frequencies, inaccessible from ground-based telescopes, so that we could test this hypothesis.

Features visible in Fig. 1 at 22 GHz are likely related to Jupiter's upper atmospheric cloud structure, seen in the IR and visible maps observed on the same day as well as previously (*12–14*). Two prominent vortex structures are visible at lower frequencies, corresponding to deeper layers, one at 38°N and the other at 19°N (planetographic latitude). These two vortices are also apparent in the visible and IR maps. They extend to depths far below the 2- to 3-bar pressure level, reaching to below the expected water condensation level. Near 10°N, within Jupiter's North Equatorial Belt (NEB), there are elongated features (known as rifts or plumes) with north-south extensions that also penetrate below the water condensation level. The zone-belt structure is evident at all MWR frequencies, as previously observed at frequencies of >8 GHz (*4*, *9*, *10*). The presence of a zone-belt structure at frequencies of <8 GHz indicates that contrasts in ammonia and/or water persist to depths well below the water

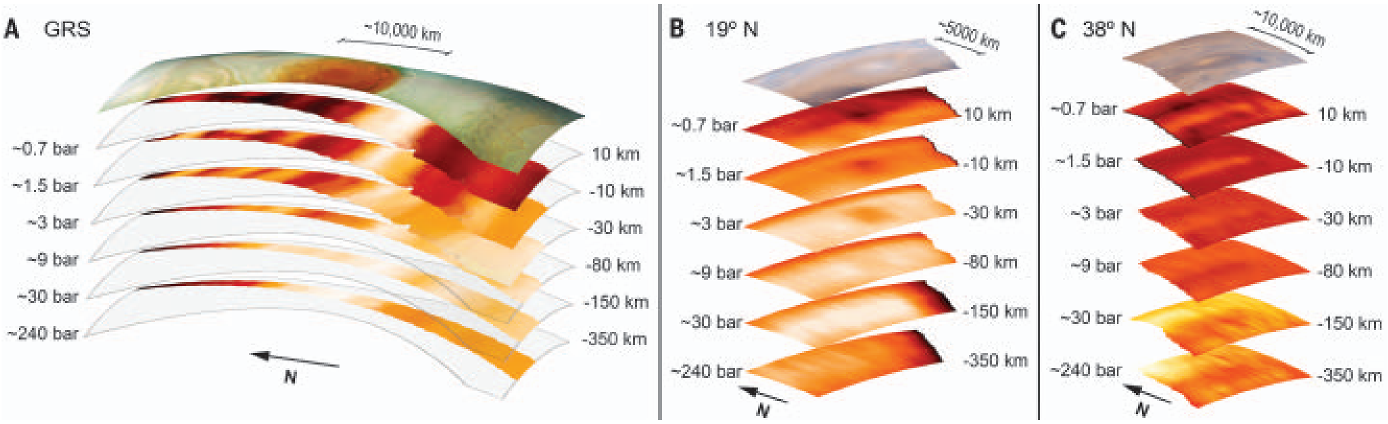


**Fig. 2. Horizontal slices of vortices at different depths.** (**A** to **C**) From top to bottom, the visible light image followed by Juno MWR data at frequencies of 22, 10, 5.2, 2.6, 1.25, and 0.6 GHz. Each slice is labeled with the corresponding pressures and depths (measured from the 1-bar pressure level). (A) The GRS in 2017. (B) 19°N storm in 2019 [top layer HST image (*13*)]. (C) 38°N storm in 2019. MWR data are colored on the same temperature scales as those in Fig. 1 for each frequency.

condensation level. Any strong vertical wind shear in these features, or zonal flows, must be accompanied by strong latitudinal variations in the atmospheric kinetic temperature ($T_\text{k}$) to maintain thermal wind balance. Gravity measurements indicate that zonal winds penetrate much deeper than that probed with the MWR (*5*, *6*), implying an absence of large latitudinal temperature variations (and therefore an absence of strong vertical wind shear). This is consistent with most of the variation in $T_\text{b}$ being due to varying amounts of ammonia and/or water that change the transparency of the atmosphere, rather than changes in $T_\text{k}$.

The feature at 38°N, 96°W (Fig. 1 and fig. S3) is an elongated cyclonic vortex, known as a barge (*12*, *15*). This vortex is evident at all MWR frequencies except at 0.6 GHz, where it is only $2\sigma$ above the background (fig. S3). At frequencies below 2.6 GHz, the barge follows the shape and extent of the signatures at higher frequencies and in the visible/IR maps, demonstrating that the deepest detection is not related to a jet stream. At pressures of <3 bar, (>5.2 GHz) the vortex has a higher $T_\text{b}$ than that of its surroundings. This reverses at pressures of >9 bar (<2.6 GHz), with the barge exhibiting a lower $T_\text{b}$ than that of its surroundings. Although our data do not directly measure wind motion, cyclonic motion is implied by the warm upper tropospheric temperatures typically observed in barges, and because vortices would be short lived if they rotated in the opposite direction to the horizontal wind shear produced by the adjacent zonal jets (supplementary text) (*15*–*19*).

At 19°N, there is a small anticyclonic vortex, just north of the NEB and within the North Tropical Zone. MWR data from 2.6 to 22 GHz show a low-$T_\text{b}$ anomaly in the upper part of this anticyclone, opposite in sign to the high-$T_\text{b}$ anomaly of the cyclone at 38°N. The low-$T_\text{b}$ anomaly fades in the 2.6 to 1.25 GHz channels, but there appears to be a ring-shaped high- $T_\text{b}$ anomaly surrounding the feature's center near 99°W at these frequencies (Fig. 1 and fig. S2). Rings associated with anticyclones have been seen at lower-pressure regions in both radio and IR observations (*20*).

In Fig. 2, nadir $T_\text{b}$ profiles of these two vortices are compared with the GRS, Jupiter's largest and possibly oldest anticyclone (*21*). A feature in the deepest MWR slice implies that the GRS base is potentially deeper than 100 bar. The top of the 19°N anticyclone is characterized by a negative ~3% $T_\text{b}$ anomaly, a characteristic similar to the GRS. However, unlike the GRS, the disturbance disappears rapidly, being barely visible at 2.6 GHz and not detected (<0.2%) in the two deepest channels. The 38°N cyclone has a different inverted structure: a positive 1 to 3% anomaly at low pressures that becomes negative at the 2.6 GHz channel and are invisible in the 0.6 GHz channel (fig. S2 and tables S1 and S2). Similar results are obtained (*8*) when comparing each vortex against its local background, both in longitude and latitude (fig. S3).

We fitted models to the data for all three vortices, applying an inversion technique to the radiance measurements to determine the vertical structure of each vortex. This method is independent of whether the observed $T_\text{b}$ anomalies are due to variations in ammonia abundance or temperature. Using the relative variation $\delta_x(P)$ to represent the pressure-dependent fluctuations arising from either ammonia abundance or kinetic temperature, we selected perturbations matching the perturbations $\Delta T_\text{b}/T_\text{b}$ seen at each MWR frequency (*8*). The results are shown in Fig. 3. Because the GRS is spatially complex and extended, we selected a single latitude to be representative of the spatially resolved northern component.

All three vortices have deep roots, extending below the water condensation level in the 5- to 10-bar region, with the GRS likely extending further than the other vortices (beyond 100 bar). Gravity measurements have independently constrained the GRS depth to <500 km (660 bar) (*22*). The best-fitting model for the 19°N vortex has a likely root depth extending to ~20 bar, with possible solutions from 8 to 80 bar. The 38°N vortex is deeper, with the best-fitting model implying a likely root depth of ~100 bar, with possible solutions from 20 to 200 bar.

The 38°N cyclone has a positive anomaly at low pressures that inverts to a negative anomaly at pressures between 3 to 6 bar and extends to pressures of 30 bar or more. This requires ammonia to be transported downward from the upper to deeper layers in Jupiter's atmosphere, possibly because of the formation of mushballs, a predicted type of multilayered ammonia-water hail (*23*, *24*). An inversion of the opposite sign is observed for the 19°N anticyclone, which is indicated by the sharp decrease in signal between 5.2 and 2.6 GHz and the overlap in the contribution functions of these frequencies (fig. S6). The transition region for this vortex is therefore located between 6 and 12 bar, which is much shallower than that inferred for the GRS. Similarly, the GRS exhibits an inversion with a brightness temperature anomaly that is negative at high frequencies and positive at low frequencies. This inversion indicates an upper part with increased volatile content (and/or lower temperature) switching to a lower part with decreased volatile content (and/or higher temperature) at pressures near ~30 bar, which is consistent with the gravity measurement (*22*).

The models of all three vortices in Fig. 3 allow a wide range of perturbations due to both ammonia and temperature variations.

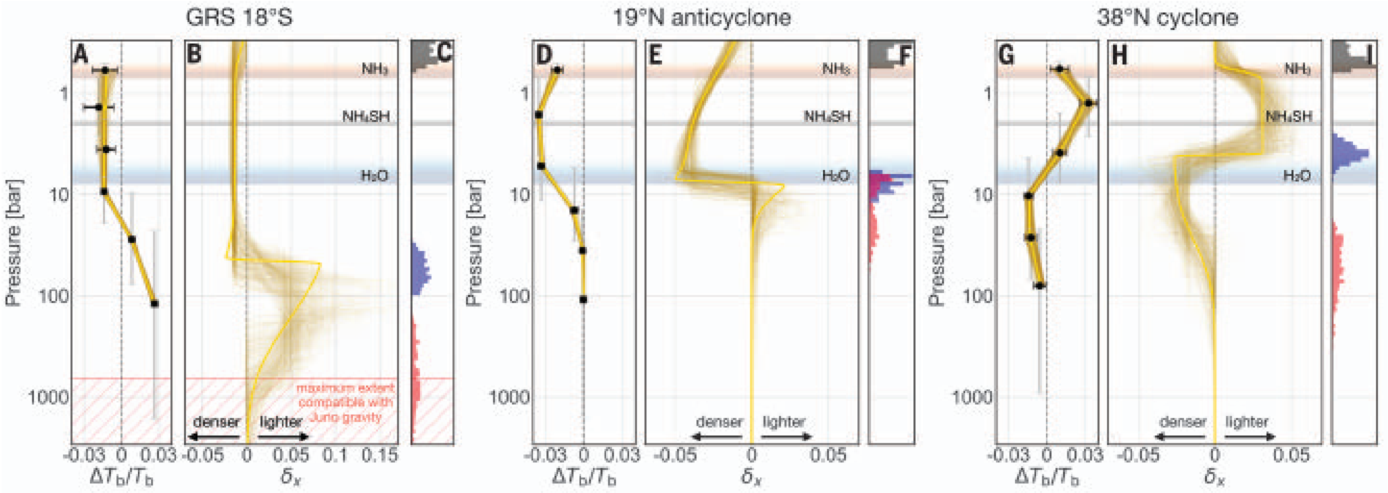


**Fig. 3. Models of Jupiter's atmospheric vortices: Observations and constraints on vortex structure.** The relative deviation between the vortex brightness temperature and the surrounding background $\Delta T_b/T_b$, the vortex perturbations $\delta_x$, and the transition regions (histograms) are shown as a function of pressure. (**A** to **C**) GRS (18°S). (**D** to **F**) 19°N anticyclone. (**G** to **I**) 38°N cyclone. (A), (D), and (G) show relative deviation between vortex $T_b$ and background. Vertical error bars indicate full width at half maximum of contribution functions (*8*). (B), (E), and (H) show relative perturbations in atmospheric opacity and/or temperature required to match MWR data. Negative values indicate higher opacity, and positive values indicate lower opacity (relative to background). In (A) to (I), models matching observations are shown in brown, and the best-fitting model is in yellow. (C), (F), and (I) show pressure histograms. Gray histograms show the lowest pressures at which the perturbation in the vortex has decayed to less than 10% of its maximum, blue histograms show the transition pressures at which the perturbation is midway between its minimum and maximum, and red histograms show the deepest pressures at which the perturbation decayed to less than 10% of its maximum. In (A) to (C), the hashed region indicates the maximum GRS extent compatible with Juno gravity measurements (*22*).

To compare the 19°N and 38°N vortices, we produced a map of the ammonia abundances (fig. S4) using data from PJ19 by using a radiative transfer model and assuming a fixed temperature profile (*9*). We found that the background atmosphere has an ammonia abundance that increases globally from the 1-bar level—where we found 130 parts per million by volume (ppmv) at 19°N and 230 ppmv at 38°N—to 350 ppmv at pressures deeper than 100 bar (fig. S5). The 19°N anticyclone has a positive +60 ppmv ammonia abundance anomaly at low pressures that inverts to a −20 ppmv anomaly near 5 bar and fades to zero at pressures deeper than 10 bar. The 38°N cyclone is characterized by a −40 ppmv anomaly at low pressures that inverts to +20 ppmv around 10 bar and fades to zero at pressures deeper than about 30 bar.

These results (Fig. 3) indicate that the midplanes of the anticyclonic vortices (19°N and the GRS) occur at different depths, very deep for the GRS but near the water condensation level for the 19°N anticyclone. The dynamics of the GRS, its circulation pattern and driving force, may be very different from other vortices in Jupiter's atmosphere. The structures of the north part of the GRS and the 19°N vortices are consistent with theoretical predictions for Jovian anticyclonic vortices, which have a dense, cold upper part over a low-density, warm lower part to maintain hydrostatic and geostrophic balance (*19*, *25*). The anticyclones are well-mixed vortices embedded in a stably stratified background medium, analogous to Earth's oceanic eddies or salt lenses (*26*–*28*). For comparison, studies of the Jovian anticyclone Oval BA indicated a transition at lower pressures, between 0.7 and 3 bar (*18*, *19*). Conversely, the 38°N vortex is more stratified than the background atmosphere. This may indicate that the thermal and compositional stratification are maintained by precipitating convection (*23*, *24*).

The deep $T_b$ anomalies we observed constrain simulations of Jupiter's vortices (*29*–*31*). The GRS and the 19°N anticyclone both have lower $T_b$ than those of the surrounding regions at high altitude and higher $T_b$ than those of the surroundings at deeper levels. These two anticyclones have a more mixed internal composition than the stratified composition in the background, whereas the 38°N cyclone shows the opposite behavior: more stratified than the background and with higher volatile concentration in the lower branch and lower volatile concentration in the upper branch. This is consistent with the presence of a deep stable layer that extends well below the water condensation level (*24*–*27*, *32*). Such a layer would result from a balance between mixing by anticyclones and mushball precipitation occurring in cyclones and embedded storms (*23*, *24*). The different characteristics and depths of the observed vortices indicate a variable extent of this stable layer as a function of latitude.

Ammonia and water have previously been considered inert tracers below the clouds so were expected to be uniformly mixed up to the relevant cloud bases, which are above the 1-bar level for ammonia and ~5- to 8-bar level for water. The mixing ratio was only expected to change within the clouds, owing to meteorological processes. The detection of individual vortex signatures at depths deeper than the water condensation level suggests the presence of small-scale dynamic processes, such as precipitation and downdrafts (*23*, *24*, *33*), at much deeper levels than expected. The sign and reversal of the $T_b$ anomalies visible in the MWR data are consistent with vortices decaying with depth (*15*). The deep roots of the GRS and other anticyclones may indicate coupling between Jupiter's interior and deep atmosphere.

## REFERENCES AND NOTES

1. H. B. Niemann *et al.*, *J. Geophys. Res.* **103** (E10), 22831–22845 (1998).
2. S. J. Bolton *et al.*, *Science* **356**, 821–825 (2017).
3. C. Li *et al.*, *Geophys. Res. Lett.* **44**, 5317–5325 (2017).
4. A. P. Ingersoll *et al.*, *Geophys. Res. Lett.* **44**, 7676–7685 (2017).
5. Y. Kaspi *et al.*, *Nature* **555**, 223–226 (2018).
6. T. Guillot *et al.*, *Nature* **555**, 227–230 (2018).
7. M. A. Janssen *et al.*, *Space Sci. Rev.* **213**, 139–185 (2017).
8. Materials and methods are available as supplementary materials.
9. F. Oyafuso *et al.*, *Earth Space Sci.* **7**, e2020EA001254 (2020).

10. I. de Pater, R. J. Sault, B. Butler, D. DeBoer, M. H. Wong, *Science* **352**, 1198–1201 (2016).
11. I. de Pater *et al*., *Icarus* **322**, 168–191 (2019).
12. R. J. Terrile, R. F. Beebe, *Science* **204**, 948–951 (1979).
13. M. H. Wong *et al*., *Astrophys. J. Suppl. Ser.* **247**, 58 (2020).
14. R. S. Giles, L. N. Fletcher, P. G. J. Irwin, *Icarus* **257**, 457–470 (2015).
15. A. R. Vasavada, A. P. Showman, *Rep. Prog. Phys.* **68**, 1935–1996 (2005).
16. A. F. Cheng *et al*., *Astron. J.* **135**, 2446–2452 (2011).
17. P. Marcus, *J. Fluid Mech.* **215**, 393–430 (1990).
18. P. S. Marcus, S. Pei, C.-H. Jiang, P. Hassanzadeh, *Phys. Rev. Lett.* **111**, 084501 (2013).
19. M. H. Wong, I. de Pater, X. Asay-Davis, P. S. Marcus, C. Y. Go, *Icarus* **215**, 211–225 (2011).
20. I. de Pater *et al*., *Icarus* **210**, 742–762 (2010).
21. B. M. Peek, *The Planet Jupiter* (Faber and Faber, 1958).
22. M. Parisi *et al*., *Science* **374**, XXXX (2021).
23. T. Guillot, D. J. Stevenson, S. K. Atreya, S. J. Bolton, H. N. Becker, *J. Geophys. Res. Planets* **125**, e2020JE006403 (2020).
24. T. Guillot *et al*., *J. Geophys. Res. Planets* **125**, e2020JE006404 (2020).
25. M. Flasar *et al*., *J. Geophys. Res.* **86** (A10), 8759–8767 (1981).
26. M. D. Prater, T. B. Sanford, *J. Phys. Oceanogr.* **24**, 1572–1586 (1994).
27. O. Aubert, M. Le Bars, P. Le Gal, P. S. Marcus, *J. Fluid Mech.* **706**, 34–45 (2012).
28. P. Hassanzadeh, P. S. Marcus, P. Le Gal, *J. Fluid Mech.* **706**, 46–57 (2012).
29. D. Lemasquerier, G. Facchini, B. Favier, M. Le Bars, *Nat. Phys.* **16**, 695–700 (2020).
30. G. Heimpel, T. Gastine, J. Wicht, *Nat. Geosci.* **9**, 19–23 (2016).
31. R. K. Yadav, M. Heimpel, J. Bloxham, *Sci. Adv.* **6**, eabb9298 (2020).
32. M. Allison, *Icarus* **83**, 282–307 (1990).
33. H. N. Becker *et al*., *Nature* **584**, 55–58 (2020).
34. M. Wong, 2017. Wide field coverage for Juno (WFCJ). Mikulski Archive for Space Telescopes (2021); doi: 10.17909/T94T1H.
35. S. Bolton *et al*., Microwave observations constrain the Depth of vortices in Jupiter's atmosphere. *Zenodo* (2021); doi: 10.5281/zenodo.5527240.

### ACKNOWLEDGMENTS

We thank T. Fujiyoshi, J. Sinclair, and T. Momary for help in acquiring the Subaru data. **Funding:** S.J.B., J.H.W., S.M.L., J.A., S.T.B., S.G., G.O., and H.N.B. were funded by the National Aeronautics and Space Administration (NASA). A portion of this research carried out at the Jet Propulsion Laboratory, California Institute of Technology, was performed under contract 80NM0018D0004 with NASA. J.L.L. was funded by the Juno Project through a subcontract with the Southwest Research Institute. T.G. was funded by the Centre National d'Etudes Spatiales. L.N.F. was supported by a Royal Society Research Fellowship and European Research Council Consolidator Grant (under the European Union's Horizon 2020 research and innovation program, grant agreement 723890). Y.K. and E.G. were sponsored by the Israeli Space Agency, the Helen Kimmel Center for Planetary Science at the Weizmann Institute of Science (WIS). M.H.W. was funded by NASA through cooperative agreement 80NSSC19M0189 and Juno Participating Scientist grant 80NSSC19K1265. M.H.W. and G.O. were supported by grants GO-14661 and GO-15665 from the Space Telescope Science Institute, which is operated by AURA under NASA contract NAS 5-26555. **Author contributions:** S.J.B. led the data analysis and interpretation of the MWR data and drafted the manuscript with input from coauthors T.G., M.H.W., Y.K., J.L.L., and S.M.L.; A.I., J.A., S.T.B., S.G., C.L., G.O., M.H.W., F.O., Z.Z., H.N.B., J.B., R.K.Y., M.A., and L.N.F. assisted with writing, interpretation, data analysis, and comparison between the HST, Gemini, and MWR data. M.H.W. and G.O. contributed to the acquisition of Gemini and HST data. J.L.L., D.S., T.G., Y.K., E.G., M.J., S.M., P.S., S.A., J.H.W., R.K.Y., and J.B. assisted with the writing, modeling, and interpretation of the MWR data. **Competing interests:** M.H.W. is also affiliated with the University of Michigan. J.L.L. is also affiliated with the Jet Propulsion Laboratory, California Institute of Technology. **Data and materials availability:** The MWR data are available from the Planetary Data System at https://pds-atmospheres.nmsu.edu/cgi-bin/getdir.pl?volume=jnomwr_1100; we used the IRDR files in directories 2019096 and 2017191. The reduced HST and Gemini data are available at (*34*). The raw HST data are available at https://archive.stsci.edu/hst/search.php under Proposal ID 15665, and raw Gemini data are at https://archive.gemini.edu/searchform under Program ID GN-2019A-Q-202. Our model results and Monte Carlo analysis code are archived at (*35*).

### SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf1015
Materials and Methods
Supplementary Text
Figs. S1 to S7
Tables S1 to S5
References (*36–41*)

6 October 2020; accepted 28 September 2021
Published online 28 October 2021
10.1126/science.abf1015

---

QUANTUM GASES

# Observation of Pauli blocking in light scattering from quantum degenerate fermions

**Amita B. Deb* and Niels Kjærgaard**

**The Pauli exclusion principle forbids indistinguishable fermions from occupying the same quantum mechanical state. The implications of this are profound; for example, it accounts for the electronic shell structure of atoms. Here, we performed measurements on the scattering of light from ultracold ensembles of atoms. For a Fermi gas in the quantum degenerate regime, we observed a marked suppression in scattering compared with a similarly prepared Bose gas. The observed decrease in fluorescence and the corresponding increase in light transmission results from Pauli blocking, where Fermi-Dirac statistics limits the number of accessible states for a scattering atom in a large Fermi sea. Our work confirms a fundamental result on the optical response of quantum gases and may contribute to cooling and thermometry schemes.**

The interaction of light with ultracold atoms is of paramount importance in emerging quantum technologies that are based on quantum gases. Understanding collective and quantum effects in the interplay between light with atomic dipoles is a highly active field in atomic physics (*1*–*4*), with its origin in (*5*), which predicted a modification of the radiative lifetime of atoms placed within a volume smaller than the cube of the wavelength of light. Another prominent example is the Purcell effect (*6*), in which the spontaneous emission rate is modified because of a change in the number of accessible states for photons. This happens, for example, in atoms in optical resonators and lies at the heart of cavity quantum electrodynamics.

In addition to playing a crucial role in the formation of the electronic structure of atoms, Pauli blocking of fermions reveals itself as anti-bunching of density-density correlations (*7*, *8*) and suppression of density fluctuations in fermionic systems (*9*–*11*). It also prevents colliding fermions from scattering sideways (*12*). For the light-matter interaction, quantum statistical effects in the spontaneous emission rate ($\gamma$) in degenerate quantum gases was first considered three decades ago (*13*). In particular, it was predicted that $\gamma$ is suppressed in a deeply degenerate Fermi gas because of Pauli blocking. As pointed out in (*14*) and (*15*), light scattering from trapped fermions establishes an analogous but experimentally more tractable scenario (Fig. 1). The possibility of suppression of light scattering has received continuous interest since the initial realization of degenerate Fermi gases (*10*, *15*–*18*), but its direct experimental detection has remained elusive.

Here, we report a direct observation of the suppression of off-resonant light scattering from a degenerate Fermi gas of $^{40}$K. We conducted experiments in three different configurations (Fig. 1). We first observed a clear reduction in fluorescence from a degenerate gas. This was complemented by transmission measurements of a probe laser beam passing through the Fermi gas. These measurements confirm that degeneracy lowers the imaginary part of the optical susceptibility of the gas. We directly compared our results for the fermionic $^{40}$K gas with those for a bosonic gas of $^{87}$Rb and observed a significant increase in light transmission for the degenerate Fermi gas relative to the Bose gas, which resulted from Pauli blocking of light scattering. The transmission through both Fermi and Bose gases in the quantum degenerate regime was observed to deviate from a Beer-Lambert law of attenuation, with light scattering being suppressed in the fermionic case and enhanced in the bosonic case. By measuring the phase shift incurred by a probe laser beam, we confirmed the prediction that degeneracy leaves the real

Department of Physics, Quantum Science Otago (QSO), and Dodd-Walls Centre for Photonic and Quantum Technologies, University of Otago, Dunedin, New Zealand.
*Corresponding author. Email: amita.deb@otago.ac.nz

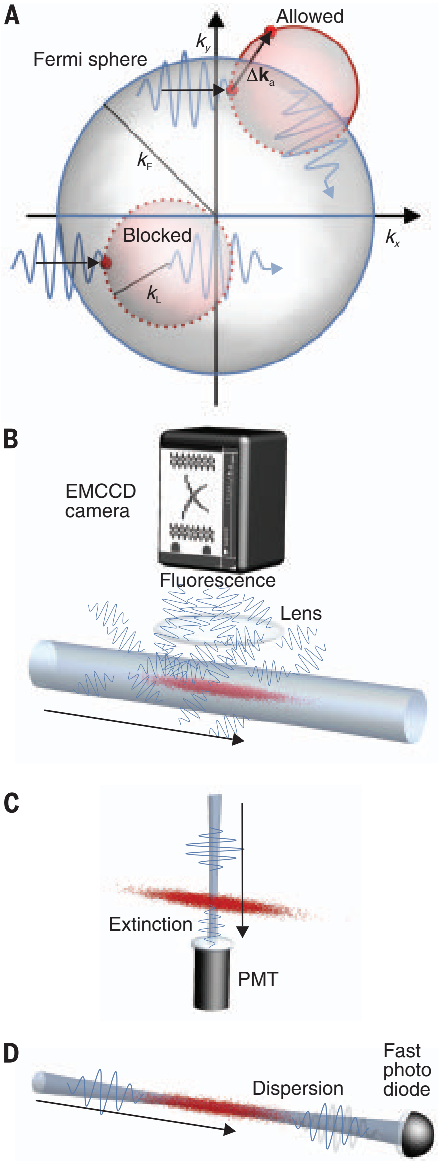


**Fig. 1. Light scattering from fermionic atoms: Schematic of the experiment.** (**A**) Fermi sphere at $T = 0$ for a trapped Fermi gas in which all states within a radius of $k_{\mathrm{F}}$ are occupied. For off-axis scattering of an incoming photon to be allowed, the associated atomic recoil $\Delta\mathbf{k}_a$ must take the scattering atom to an unoccupied state outside the Fermi sphere. (**B**) Laser-induced fluorescence from a uniformly illuminated spheroidal cloud of atoms recorded in the transverse direction using an EMCCD camera. (**C**) Extinction of a Gaussian laser beam propagating radially through the center of a spheroidal cloud of atoms measured with a PMT. (**D**) Dispersive phase shift incurred by a Gaussian laser beam propagating coaxially through a spheroidal cloud of atoms measured on a fast photodiode using a heterodyne interferometric scheme.

part of the susceptibility of a Fermi gas largely unaffected.

The physical mechanism behind Pauli blocking of light scattering is illustrated in Fig. 1A, where we show a momentum space representation for a harmonically trapped Fermi gas with $N$ atoms at temperature $T = 0$. The atoms occupy all states within the Fermi sphere of radius $k_{\mathrm{F}} = \sqrt{2mE_{\mathrm{F}}}/\hbar$, where $\hbar$ is the reduced Planck constant, $m$ is the atomic mass, and $E_{\mathrm{F}} = \hbar\bar{\omega}(6N)^{1/3}$ is the Fermi energy when the geometric mean of the trapping frequencies is $\bar{\omega}$. Upon scattering a photon from an incoming beam of light with a wavelength $\lambda$, an atom will incur a change in wave number of $|\Delta\mathbf{k}_a| = 2k_{\mathrm{L}}\sin(\theta/2)$, where $\theta$ is the angle between the directions of the incident and the scattered photon and $k_{\mathrm{L}} = 2\pi/\lambda$. For $k_{\mathrm{F}} > k_{\mathrm{L}}$, recoil events connecting an atom to an occupied state inside the Fermi sphere would be forbidden by the Pauli exclusion principle. To observe this blocking effect clearly in an experiment, the photon recoil energy $E_{\mathrm{R}} = \hbar^2k_{\mathrm{L}}^2/2m$ should preferably be much less than $E_{\mathrm{F}}$, and it is therefore advantageous to use relatively heavy fermionic atoms such as $^{40}$K and a large number of atoms, as in our case. The number of vacancies inside the Fermi sphere will increase with $T$, so effective Pauli blocking also requires the sample temperature to be low compared with the Fermi temperature $T_{\mathrm{F}} = E_{\mathrm{F}}/k_{\mathrm{B}}$, where $k_{\mathrm{B}}$ is Boltzmann's constant.

In our experiment, we produced quantum degenerate Fermi gases of $^{40}$K in the ground hyperfine state $|4S_{1/2}; F = 9/2, m_{\mathrm{F}} = 9/2\rangle$ harmonically confined by an Ioffe-Pritchard type of magnetic trap. The trap has radial and axial trapping frequencies of $\omega_{\mathrm{r}} = 2\pi \times 260$ Hz and $\omega_{\mathrm{a}} = 2\pi \times 30$ Hz, respectively, corresponding to $\bar{\omega} = 2\pi \times 127$ Hz. By using a sympathetic cooling procedure with $^{87}$Rb atoms as the coolant, we were able to produce characteristic $^{40}$K samples with $N = 2.2 \times 10^6$ atoms at temperatures ~420 nK, corresponding to $T/T_{\mathrm{F}} \lesssim 0.3$. For benchmarking, we could also produce ultracold bosonic samples of $^{87}$Rb in the $|5S_{1/2}; F = 2, m_{\mathrm{F}} = 2\rangle$ hyperfine state, for which the radial and axial trapping frequencies are $\omega_{\mathrm{r}} = 2\pi \times 170$ Hz and $\omega_{\mathrm{a}} = 2\pi \times 20$ Hz. For characteristic samples of $N = 2.2 \times 10^6$, the critical temperature for Bose-Einstein condensation (BEC) $T_{\mathrm{c}} = \hbar\bar{\omega}(N/1.2)^{1/3}$ is ~500 nK. Our experimental setup incorporated three in situ diagnostics shown schematically in Fig. 1, B to D. In addition, we could release a sample from the trap to perform time-of-flight absorption imaging. The number and temperature of a fermionic sample could be independently controlled by adjusting the initial loading of the trap and the final microwave frequency for forced evaporation of the coolant atoms.

In Fig. 2, we show the temperature dependence of atomic fluorescence from a $^{40}$K gas with $N = 2.1 \times 10^6$ as it enters the quantum degenerate regime. The measurements were conducted by illuminating the gas with a wide-probe laser beam (4.8-mm spot-size radius) for 3 ms and acquiring fluorescence images (Fig. 2A) with an electron-multiplying charge-coupled device (EMCCD) (Fig. 1B). The photon detection efficiency at 90° is 0.6%, including the quantum efficiency of the EMCCD camera. The probe light was $\sigma^+$ polarized and detuned by $\Delta = -500\,\Gamma$ from resonance of the $|4S_{1/2}; F = 9/2, m_{\mathrm{F}} = 9/2\rangle \leftrightarrow |4P_{3/2}; F = 11/2, m_{\mathrm{F}} = 11/2\rangle$ optical cycling transition, where $\Gamma = 2\pi \times 6$ MHz is the natural linewidth. We used a probe beam power corresponding to an on-resonance saturation parameter of ~3.5, which for the detuned probe gives a photon scattering rate $\mathcal{R}_{\mathrm{sc}} = 2\pi \times 10$ Hz for an independent atom. An atom scatters ~0.2 photons on average during the process and, as a result, the perturbation caused by the probe pulse on the temperature and the Fermi sea occupation of the ensemble can be considered negligible, which we verified using absorption imaging (*19*). In Fig. 2, Pauli blocking of light scattering manifests as a reduction in the number of photons collected by the EMCCD as $T$ becomes lower. Figure 2B shows the total number of collected fluorescence photons as a function of $T/T_{\mathrm{F}}$; we found excellent agreement with the prediction from the semiclassical model in (*15*), which gives the suppression for a given angle using a Fermi golden rule approach (*19*).

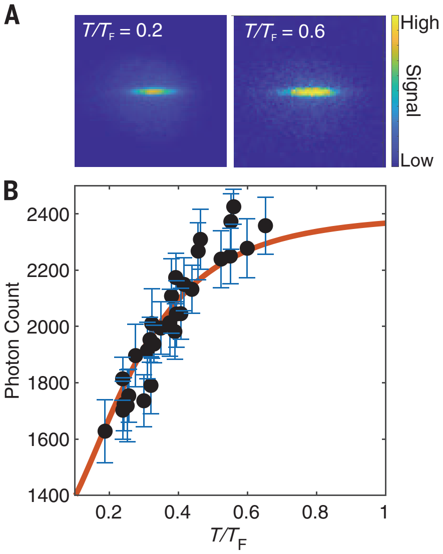


**Fig. 2. Laser-induced fluorescence from trapped fermionic $^{40}$K clouds.** (**A**) Examples of in-trap fluorescence images for clouds with $N = 2.1 \times 10^6$ atoms cooled to $T/T_{\mathrm{F}} = 0.2$ (left) and $T/T_{\mathrm{F}} = 0.65$ (right). The signal from the colder and more degenerate cloud is suppressed. Each image is an average of 20 experimental realizations. (**B**) Number of photons collected by the EMCCD as a function of $T/T_{\mathrm{F}}$, with the first and last data point corresponding to the images in (A). The line shows the prediction by the suppression model from (*15*) with no free parameters.

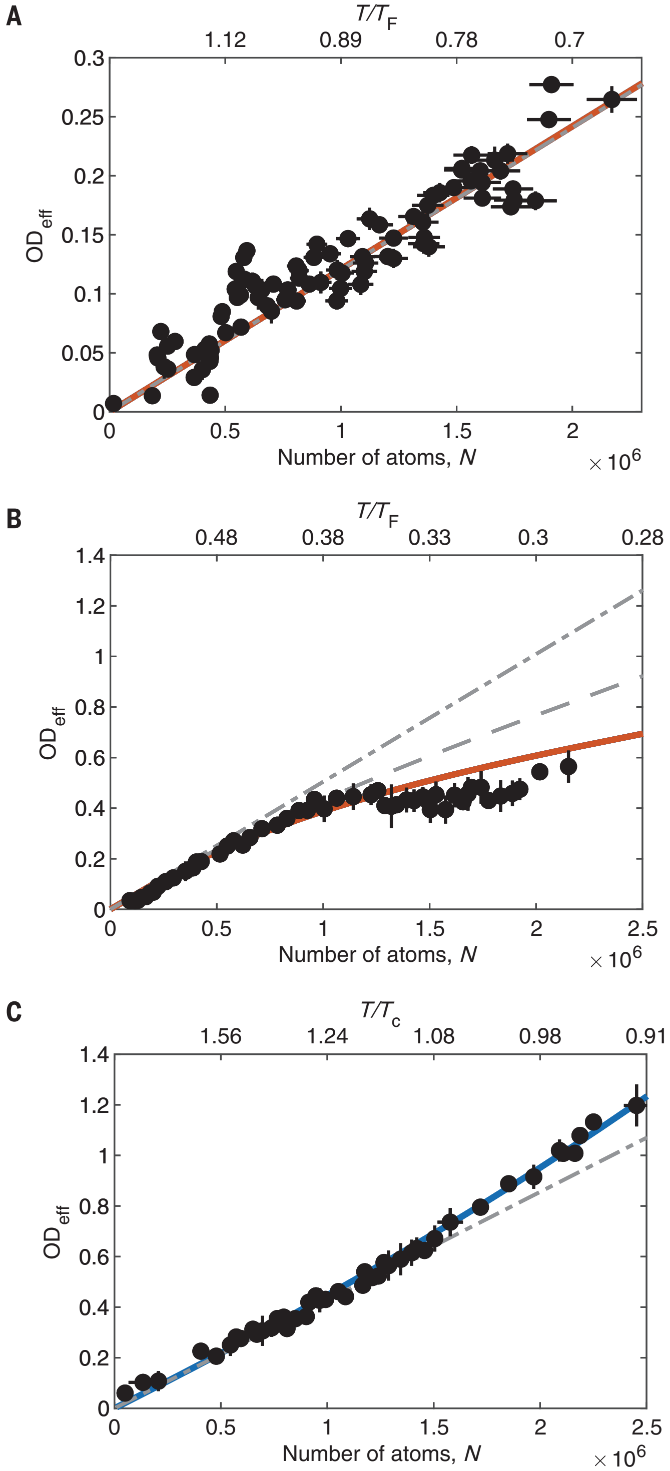


**Fig. 3. Light extinction by trapped atomic gases.** Shown is the effective optical depth as a function of atom number, measured in a radial direction through the center of an atomic cloud (see Fig. 1C). (**A** and **B**) Fermionic $^{40}$K samples at $T$ = 980 nK for a range of atom numbers corresponding to limited quantum degeneracy (A) and samples at $T$ = 420 nK exploring a region of significant degeneracy (B). In both (A) and (B) $T/T_F$ is shown on the top axis. (**C**) Bosonic $^{87}$Rb at $T$ = 480 nK. The top axis gives $T/T_c$ corresponding to the atom number $N$. In (A) to (C), the solid lines (red for fermions and blue for bosons) show the result of the MCD model, and the dashed-dotted lines show the prediction for a classical Maxwell-Boltzmann gas (*19*). The dashed line in (B) shows the prediction for a classical gas with a density profile modified by Fermi pressure.

A reduction in light scattering from a sample is accompanied by an increase in transmission of a probe beam. To measure this effect with a high signal-to-noise ratio while keeping the multiple scattering of photons low, an optical density on the order of one is optimal (*15*). This required us to operate much closer to the atomic resonance than for the above fluorescence measurements. We used a probe beam that propagates along the short axis of the cloud (Fig. 1C) and is focused to a $1/e^2$ waist size of 19 μm. The probe light is polarized perpendicularly to the long axis of the cloud and has a frequency detuning of $\Delta = -9\Gamma$. The light after propagation through the atoms was captured by an optical system with an effective numerical aperture of ~0.165 and was detected using a photomultiplier tube (PMT). In this manner, we were able to preferentially sample the response of the central part of the atomic distribution, which we verified by measuring optical depths of thermal atomic ensembles with a known in-trap density distribution (*19*). The probe beam has a power of ~2 nW and is pulsed for a duration of 90 μs. Under these conditions, an independent atom would scatter around 0.2 photon during the probe pulse.

We determined the ratio $P_t/P_0$ between the light transmitted through $^{40}$K samples with and without the atoms as measured on the PMT. Figure 3A shows $\mathrm{OD}_{\mathrm{eff}} = -\ln(P_t/P_0)$ as a function of atom number for a 980 (±50) nK sample. The exponentially decaying transmission profile is well described by the Beer-Lambert extinction law. By contrast, Fig. 3B presents the results of measurements on samples with a temperature of 420 (±40) nK over the same range of atom numbers, for which we observed a clear departure from Beer-Lambert behavior as the number of atoms increased. The crucial difference between the scenarios in Fig. 3, A and B, is the degree of quantum degeneracy, as given by $T/T_F$, with the colder gas exploring a more deeply degenerate regime as quantified by $T/T_F$ (see top axes). In Fig. 3C, we show measurements on samples of bosonic $^{87}$Rb atoms at temperature 480 (±50) nK under similar probing conditions. Compared with this bosonic case, the transmission of light through fermionic atomic clouds presented in Fig. 3B is substantially enhanced for higher atom numbers, where the degeneracy is appreciable.

We modeled our transmission measurements by applying a theory introduced by Morice, Castin, and Dalibard (MCD) in (*20*), which describes the coherent optical response of a gas of atoms with complex polarizability α. The MCD theory shows that in the regime of low atomic density $\rho \ll k_L^3$, the spatial correlations resulting from the quantum statistics describing the atoms near or below quantum degeneracy modifies the optical susceptibility $\chi_0 = \rho\alpha/\varepsilon_0$ of the gas according to (*21*):

$$\chi = \chi_0 \frac{1}{1+\mathcal{A}} \tag{1}$$

where

$$\mathcal{A} = \underbrace{-\frac{1}{3}\frac{\rho\alpha}{\varepsilon_0}}_{\text{Lorentz−Lorenz correction}} \underbrace{-\frac{\rho\alpha}{\varepsilon_0}\int d^3\mathbf{r}\, g_{xx}(\mathbf{r})\phi_\rho(\mathbf{r})e^{-i\mathbf{k}_L\cdot\mathbf{r}}}_{\text{Pair correlation correction}} \tag{2}$$

and $\phi_\rho(\mathbf{r})$ is the atomic pair correlation function at an atomic density ρ, reflecting the spatial bunching (bosons) or antibunching (fermions), and $g_{xx}(\mathbf{r})$ is the component of a tensor related to the angular pattern of the field radiated by an oscillating dipole along the interatomic axis (*19*). At the maximum atomic densities used in our experiments, we obtained $\rho/k_L^3 \lesssim 0.1$, and the Lorentz-Lorenz correction term in Eq. 2 is negligible for our detunings. For the quantum degenerate regime reached in our experiments, the pair correlation correction, however, becomes substantial.

To apply the MCD description to our system, we used Eq. 1 to predict $\chi$ across a sample at a given temperature using the known polylogarithmic density distributions and the pair correlations functions (fig. S1) appropriate for either Fermi and Bose gases (*19*). Convoluting the imaginary part of the corrected $\chi$ with the Gaussian probe beam profile, we obtained an overall correction factor for $OD_{eff}$. The solid lines in Fig. 3 present the predicted $OD_{eff}$ from the MCD model as a function of $N$, and the dashed-dotted lines show the predicted $OD_{eff}$ for a classical gas without the corrections applied. The latter is linear in atomic number, in accordance with a Beer-Lambert law dependence, and the low-degeneracy Fermi gas in Fig. 3A matches this behavior well. By contrast, the data in Fig. 3B, in which the degeneracy becomes significant, clearly deviates from a Beer-Lambert law displaying a reduction in $OD_{eff}$, which is captured well by the MCD model. Part of this reduction is caused by the transverse density distribution becoming non-Gaussian and saturating at the center because of Fermi pressure, limiting the number of atoms inside the volume of the Gaussian probe beam. The dashed line shows the predicted $OD_{eff}$ with the Fermi pressure effect on the density distribution included for the otherwise classical gas, and it is clear that the MCD model, which includes the correlation correction, the source of Pauli blocking, provides a much better description of the data. Finally, in Fig. 3C, we see that the deviation in $OD_{eff}$ from linearity for our bosonic benchmark becomes significant when $T$ approaches $T_c$ from above and is also captured well by the MCD model. This enhancement in light scattering across $T_c$ is expected from the MCD theory because of a stronger bunching of the pair correlation function at the length scale $\lambda$. This results in significantly increased density fluctuations on this length scale, which from a fluctuation-dissipation scenario (*10*), suggests increased light scattering in this regime. Indeed, a sharp onset in density fluctuations across the BEC phase transition has recently been observed directly (*22*, *23*) and by increased recoil heating (*24*).

Whereas the imaginary part of $\chi$ for fermions is substantially influenced by quantum degeneracy, the MCD model predicts that its real part is left largely unaffected (fig. S2). To map the dispersive response of the atoms, we used frequency modulation spectroscopy (FMS) (*25*, *26*), probing our trapped Fermi gases with a light beam containing an electro-optically generated frequency triplet (a carrier and two sidebands). Only one sideband is significantly affected by the atoms because of its proximity to the atomic resonance, whereas the far-detuned carrier acts as a reference arm for a heterodyne interferometer (*19*). The $\sigma^+$-polarized probe light beam propagates along the axial direction of the atomic cloud (Fig. 1D) and is focused to a spot-size radius of 25 μm. The FMS signal is sensitive to both the extinction and the dispersive phase shift of the light, but by using a large detuning $\Delta = -420\Gamma$, we were able to make the dispersive part dominant. Figure 4 shows the FMS signal as a function of atom number for $^{40}$K samples at 530 nK, to which we fit a straight line through the origin. The clear direct proportionality relation between the phase shift of light and the atom number shows that the dispersive response of an atomic medium is practically unmodified by Fermi degeneracy. In the Fermi golden rule scenario, the Pauli exclusion principle does not prohibit a process in which an atom returns to its original motional state after scattering a photon. Because it is such elastic scattering that is responsible for the phase shift of light, one does not expect any Pauli blocking. In the MCD model, a small energy shift in the resonance position is predicted as a result of quantum degeneracy. This cooperative Lamb shift leads to a very small modification of the real part of $\chi$, and the MCD model is thus consistent with causality and the Kramers-Kronig relations (*27*).

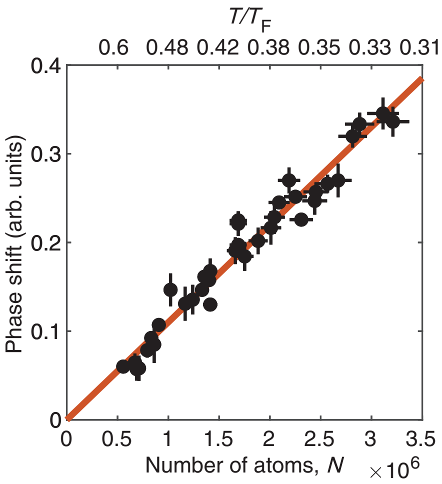


**Fig. 4. Phase shift of light in trapped fermionic $^{40}$K gases.** Phase shift incurred by a probe laser beam propagating along the axis of a spheroidal atomic cloud (see Fig. 1D) measured using frequency modulation spectroscopy. The solid line is a fit to a directly proportional relationship between the phase shift and number of atoms in the cloud.

Our observations provide insights into a fundamental theoretical problem concerning the optical response of quantum gases and pave the way for cooling and thermometry schemes for fermions deep below the Fermi temperature and quantum non-demolition measurements with reduced dissipation (*28*).

## REFERENCES AND NOTES

1. R. J. Bettles, S. A. Gardiner, C. S. Adams, *Phys. Rev. Lett.* **116**, 103602 (2016).
2. J. Pellegrino *et al.*, *Phys. Rev. Lett.* **113**, 133602 (2014).
3. S. L. Bromley *et al.*, *Nat. Commun.* **7**, 11039 (2016).
4. J. Rui *et al.*, *Nature* **583**, 369–374 (2020).
5. R. H. Dicke, *Phys. Rev.* **93**, 99–110 (1954).
6. E. M. Purcell, H. C. Torrey, R. V. Pound, *Phys. Rev.* **69**, 37 (1946).
7. W. D. Oliver, J. Kim, R. C. Liu, Y. Yamamoto, *Science* **284**, 299–301 (1999).
8. M. Henny *et al.*, *Science* **284**, 296–298 (1999).
9. T. Müller *et al.*, *Phys. Rev. Lett.* **105**, 040401 (2010).
10. C. Sanner *et al.*, *Phys. Rev. Lett.* **105**, 040402 (2010).
11. A. Omran *et al.*, *Phys. Rev. Lett.* **115**, 263001 (2015).
12. R. Thomas *et al.*, *Nat. Commun.* **7**, 12069 (2016).
13. K. Helmerson, M. Xiao, D. Pritchard, in Proceedings of the International Quantum Electronics Conference, 21–25 May 1990, Anaheim, CA, USA (Optical Society of America, 1990), p. QTHH4.
14. T. Busch, J. Anglin, J. Cirac, P. Zoller, *Europhys. Lett.* **44**, 1–6 (1998).
15. B. Shuve, J. Thywissen, *J. Phys. At. Mol. Opt. Phys.* **43**, 015301 (2009).
16. B. DeMarco, D. Jin, *Phys. Rev. A* **58**, R4267–R4270 (1998).
17. B. DeMarco, D. S. Jin, *Science* **285**, 1703–1706 (1999).
18. J. Ruostekoski, C. J. Foot, A. B. Deb, *Phys. Rev. Lett.* **103**, 170404 (2009).
19. Materials and methods are available as supplementary materials.
20. O. Morice, Y. Castin, J. Dalibard, *Phys. Rev. A* **51**, 3896–3901 (1995).
21. O. Morice, "Atomes refroidis par laser: du refroidissement sub-recul à la recherche d'effets quantiques collectifs," thesis, Université Pierre et Marie Curie, Paris (1995).
22. M. B. Christensen *et al.*, *Phys. Rev. Lett.* **126**, 153601 (2021).
23. M. A. Kristensen *et al.*, *Phys. Rev. Lett.* **122**, 163601 (2019).
24. P. C. Bons, R. de Haas, D. de Jong, A. Groot, P. van der Straten, *Phys. Rev. Lett.* **116**, 173602 (2016).
25. G. C. Bjorklund, *Opt. Lett.* **5**, 15 (1980).
26. A. B. Deb, B. J. Sawyer, N. Kjærgaard, *Phys. Rev. A* **88**, 063607 (2013).
27. J. D. Jackson, *Classical Electrodynamics* (Wiley, ed. 2, 1975).
28. R. Onofrio, *Phys. Uspekhi* **59**, 1129–1153 (2016).
29. Data and code for: A. B. Deb, N. Kjærgaard, Observation of Pauli blocking in light scattering from quantum degenerate fermions. Zenodo (2021); https://zenodo.org/record/5506594#.YUBhTi2B3wd.
30. Y. Castin, J. I. Cirac, M. Lewenstein, *Phys. Rev. Lett.* **80**, 5305–5308 (1998).
31. A. Urvoy, Z. Vendeiro, J. Ramette, A. Adiyatullin, V. Vuletić, *Phys. Rev. Lett.* **122**, 203202 (2019).
32. S. Inouye *et al.*, *Science* **285**, 571–574 (1999).
33. A. B. Deb, J. Chung, N. Kjærgaard, *New J. Phys.* **22**, 073017 (2020).

## ACKNOWLEDGMENTS

We thank C. Gardiner, S. Hoinka, J. Ruostekoski, and C. Vale for discussions; B. Sawyer and R. Thomas for developing the dispersive probing setup and the high-atom-number Fermi gas setup, respectively, and I. Kinski for manufacturing isotopically enriched $^{40}$K dispensers. **Funding:** This work was supported by the Marsden Fund of New Zealand (contracts UOO1923 and UOO1729). **Author contributions:** The experimental work and data analysis were done by A.B.D. Both authors contributed to interpreting the results, making the figures, and writing the manuscript based on a first draft written by A.B.D. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data and code are archived on Zenodo (*29*).

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abh3470
Materials and Methods
Supplementary Text
Figs. S1 to S3
Table S1
References (*30*–*33*)

3 March 2021; resubmitted 11 June 2021
Accepted 1 October 2021
10.1126/science.abh3470

### QUANTUM GASES

# Pauli blocking of light scattering in degenerate fermions

Yair Margalit[1,2]*, Yu-Kun Lu[1,2], Furkan Çağrı Top[1,2], Wolfgang Ketterle[1,2]

**Pauli blocking of spontaneous emission is responsible for the stability of atoms. Electrons cannot decay to lower-lying internal states that are already occupied. Pauli blocking also occurs when free atoms scatter light elastically (Rayleigh scattering) and the final external momentum states are already populated. This was predicted more than 30 years ago but is challenging to realize experimentally. Here, we report on Pauli blocking of light scattering in a dense quantum-degenerate Fermi gas of ultracold lithium atoms. When the Fermi momentum is larger than the photon recoil, most final momentum states are within the Fermi surface. At low temperature, we find that light scattered even at large angles is suppressed by 37% compared with higher temperatures, where atoms scatter at the single-atom Rayleigh scattering rate.**

Suppression of light scattering in ultracold Fermi gases has been predicted in works dating back to 1990 (*1*–*8*). The basic phenomenon is shown in Fig. 1. Light scattering between photon states with wave vectors $k_\mathrm{i}$ and $k_\mathrm{f}$ transfers momentum $\hbar q = \hbar(k_\mathrm{i} - k_\mathrm{f}) = 2\hbar k \sin\theta/2$, where $\hbar k_\mathrm{i}$ is the initial photon momentum, $\hbar k_\mathrm{f}$ is the final photon momentum, and $\theta$ is the scattering angle. When the Fermi momentum $\hbar k_\mathrm{F}$ of a zero-temperature Fermi gas is larger than the momentum transfer $\hbar q$, light scattering is strongly suppressed and can occur only near the Fermi surface, whereas for temperatures $T \geq T_\mathrm{F}$, the scattering rate per atom approaches the independent atom limit. This smooth transition versus temperature has been theoretically studied, including by averaging over the inhomogeneous density distribution of a harmonically trapped atom cloud (*6*).

Experiments on ultracold atoms have deepened our understanding of basic physical phenomena by realizing paradigmatic idealized situations where the phenomenon is observed in its most direct and transparent form. These realizations then become building blocks for more-complex systems. Examples include the realization of Bose-Einstein condensation (BEC), the BEC–Bardeen-Cooper-Schrieffer (BCS) crossover in fermions, band structure phenomena of noninteracting atoms in optical lattices, and Mott insulators in optical lattices (*9*). Here, we study, in a highly idealized situation, how ultracold fermions scatter light and observe the suppression of light scattering caused by Pauli blocking in a degenerate Fermi gas. Recently, we have been able to prepare ultracold fermions at very high densities (up to $n = 3 \times 10^{15}\ \mathrm{cm}^{-3}$) (*10*), where the Fermi energy is 50 times as high as that of the photon recoil energy $\hbar^2 k_\mathrm{i}^2/2m$ of 73.9 kHz (where $m$ is the lithium atomic mass). Using this sample, we have now performed light-scattering experiments in the simplest possible limit, at detunings $\Delta$ from the atomic resonance of more than 100 GHz, or 17,000 linewidths $\Gamma$. Therefore, despite high atomic densities $n \approx 1.2/\lambda^3$ and high resonant optical densities $6\pi n \lambda^2 l \approx 44{,}000$ (where $\lambda = 1/k$ and $l$ is the length of the atom cloud), we realize the limit where both the absorptive and dispersive parts of the index of refraction are negligible. In general, optical properties become complicated in the regime of high densities thanks to strong Lorentz-Lorentz corrections (*11*) and dipolar interactions between the atoms (*12*). These corrections are often expanded in the parameter $n\alpha$, where $\alpha$ is the atomic polarizability, given for a two-level atom by $\alpha = 6\pi\lambda^3\Gamma/(\Delta + i\Gamma)$. At our detunings, the parameter $n\alpha \approx 1/800$, and those corrections are negligible. Also, at detunings larger than the fine-structure splitting of 10 GHz, optical pumping to other hyperfine states is suppressed. At 100-GHz detuning, the branching ratio is <1% for any polarization of light, so no special cycling transition is needed. We use rather weak and long laser pulses with a Rayleigh scattering rate around or below 1 photon per atom during 1 ms to stay far away from nonlinear collective light scattering (*13*).

Ultracold lithium clouds were prepared as in our previous work (*10*). In short, $^{23}$Na and $^{6}$Li atoms are first laser cooled and then transferred into a quadrupole magnetic trap with an optical plug (*14*). Forced microwave evaporation of the Na atoms (*15*) sympathetically cools the lithium atoms. The lithium atoms are transferred into a single-beam 1064-nm optical dipole trap with variable spot size and power, which controls the trap volume and densities. A partially nonadiabatic radio-frequency (RF) Landau-Zener sweep transfers the majority of the atoms to the collisionally stable lowest Zeeman state $|F = 1/2, m_\mathrm{F} = 1/2\rangle \equiv |1\rangle$ while keeping ~7% in the original state $|3/2, 3/2\rangle \equiv |6\rangle$. This creates a spin mixture with s-wave interactions, which allows for efficient evaporative cooling into quantum degeneracy. Decreasing the spot size of the trapping beam creates a tighter trap with frequencies of $\omega_r/2\pi = 34$ kHz and $\omega_z/2\pi = 770$ kHz in the radial and axial directions, respectively. The atoms are exposed to a final stage of evaporative cooling by tilting the trapping potential with a magnetic gradient for 1.5 s. A typical sample contains $N \simeq 8 \times 10^5$ $^{6}$Li atoms at $T/T_\mathrm{F} \simeq 0.2$, with a Fermi temperature of $T_\mathrm{F} = \hbar\left(\omega_r^2\omega_z 6N\right)^{1/3} = 70\ \mu\mathrm{K}$. This corresponds to a density of $\sim 1 \times 10^{15}\ \mathrm{cm}^{-3}$ and an on-resonance optical density of ~25,000. We can produce even higher densities of up to $3 \times 10^{15}\ \mathrm{cm}^{-3}$ and Fermi energies of 190 μK, but they suffer from three-body losses and associated heating [which occur even in a spin-polarized sample (*10*)]. As the final step before the light-scattering experiment, the majority of the

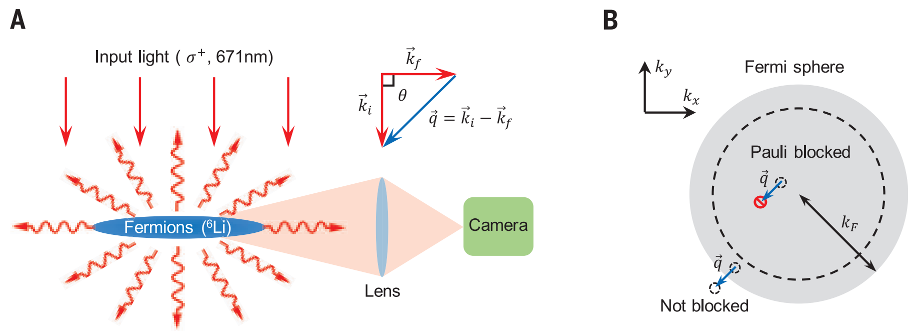


**Fig. 1. Schematic of the experiment.** (**A**) Degenerate fermionic lithium atoms are confined in an optical dipole trap and illuminated with a laser beam. Scattered photons impart momentum transfers of $\hbar q$ to the atoms and are detected at a scattering angle of 90°. (**B**) Mechanism of Pauli blocking in degenerate Fermi gases. At temperature $T = 0$, the atoms occupy a Fermi sphere in momentum space with radius $\hbar k_\mathrm{F}$. For $q \ll k_\mathrm{F}$, atoms can scatter light only from the outer shell of the Fermi sphere (of width $\hbar q$), where they can reach an unoccupied final momentum state. No scattering is possible for atoms within the dashed circle.

---

[1] Research Laboratory of Electronics, MIT-Harvard Center for Ultracold Atoms, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2] Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: margalya@mit.edu

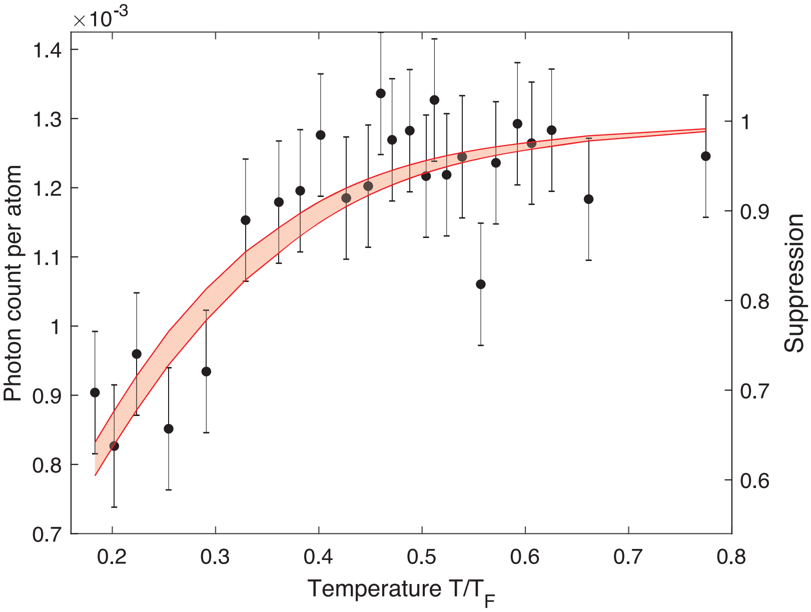


**Fig. 2. Pauli blocking of light scattering.** Photon count per atom as a function of the cloud's temperature, observed at a scattering angle of 90°. At a low temperature, the scattered light is reduced by 35% with respect to the unblocked case. The cloud is heated by turning the optical dipole trap off and on for a variable duration. The probe light is pulsed on for 25 ms at an intensity of $7.0 \times 10^2$ m W cm$^{-2}$ and is detuned 100 GHz below the atomic resonance (located at 671 nm). The two solid red lines (enclosing the red-shaded region) show the theoretical prediction of the Pauli suppression factor (right *y* axis) for the optical dipole trap potential (see the text). The difference between the solid lines represents the uncertainty in the number of atoms (40%). There are no free-fitting parameters apart from the overall vertical scale relative to the data. The error bars in all of the figures are purely statistical and reflect one standard error of the mean. Data points here are each averaged over 12 samples. The constant error bars shown are averaged over the whole dataset and reflect the best estimate for the statistical uncertainty. The uncertainties are dominated by camera-read noise for the 81 pixels within the region of interest and fluctuations of light from the trapping laser, which propagates into the camera and could not be completely suppressed by filters.

atoms are transferred by the same RF Landau-Zener sweep back to state 6, leaving ≤10% of the total number in state 1 to ensure thermalization. State 6 has a cycling transition that matters only at smaller detunings. The number of atoms in the trap is measured using standard time-of-flight absorption imaging with an estimated uncertainty of up to 40%.

After cooling to the lowest temperature, the sample is heated either by strongly modulating the trapping potential or by the scattering of light. For the observation of suppressed light scattering, we typically scatter 0.4 photons per atom during 25 ms and collect the fluorescence at a right angle with an imaging system with a collection efficiency of 0.31% (calibrated by on-resonant light scattering of a laser beam with known power and beam waist). This scattering corresponds to a recoil heating of ~4.6% of the original temperature. Given that we have reached quantum degeneracy far above the recoil temperature, scattering even a few photons does not cause substantial heating. The size of the laser beam is chosen to be much larger than the cloud size (14 times as large axially and 33 times as large radially) so that the intensity inhomogeneity across the atom cloud is negligible (*13*).

Figure 2 shows the main result of this paper—the suppression of light scattering by a degenerate Fermi gas. For each shot, the number of scattered photons is recorded, and the number of atoms and the temperature are obtained by time-of-flight absorption imaging. Thus, we measure the number of scattered photons per atom as a function of the cloud's temperature, and we observe that, in the degenerate regime ($T/T_\mathrm{F} \approx 0.2$), the atoms scatter ~35% less light thanks to Pauli blocking compared with the unblocked case. Results are limited to $T/T_\mathrm{F} \leq 0.8$ to reduce systematic errors, such as atom loss by spilling caused by the finite trap depth. Our observations are in good agreement with theoretical calculations for a trapped cloud of atoms (*6*). The theoretical model extends the treatment of (*6*) to the Gaussian potential of the optical dipole trap (*13*). Compared with a harmonic trap, the anharmonicites lead to smaller Pauli blocking (for our ratio of trap depth to $E_\mathrm{F}$ of ≈5.7), and the light-scattering rate approaches the nondegenerate limit to within 2% at $T/T_\mathrm{F} = 0.7$. The lowest temperatures are measured from the shape of the degenerate cloud [by fitting to a polylog function (*16*)], directly providing the fugacity or $T/T_\mathrm{F}$ without any correction parameters. For higher temperatures, when the cloud shape becomes Gaussian, this direct method fails, and we instead determine the temperature from Gaussian fits to the wings of the cloud and $k_\mathrm{B}T_\mathrm{F} = \hbar\left(\omega_r^2\omega_z 6N\right)^{1/3}$ from the number count $N$, where $k_\mathrm{B}$ is the Boltzmann constant. For a broad range of intermediate temperatures, the two methods agree to within a correction factor that accounts for experimental drifts between the data run and when the $N$ count and trap frequencies were carefully calibrated. This correction factor for Fermi energies (which varied for different data runs between 1.1 and 1.2) affects only the horizontal scale and not the Pauli suppression. For the comparison of experimental data with the theoretical curve in Fig. 2, the only adjustable parameter is the normalization of the high-temperature photon signal to 1. This normalization is equivalent to a calibration of the excitation and detection efficiencies for the exact experimental conditions in which the data were taken.

Because light scattering heats up the cloud by photon recoil heating, Pauli suppression can be observed only for sufficiently short or weak laser pulses. This is demonstrated in Fig. 3, where we study the number of photons scattered per atom as a function of probe laser power for initially degenerate and nondegenerate clouds. For a nondegenerate cloud ($T/T_\mathrm{F} \approx 0.7$), the photon scattering signal is linear in laser power for the whole range of powers studied, with a slope of 2.3 ± 0.09 [1/W] (Fig. 3B). For a degenerate cloud, there is an initial linear regime for low power (which was used for Fig. 2), but after scattering several photons per atom, the temperature increases owing to recoil heating, and the slope increases as a consequence of the gradual elimination of Pauli blocking. The blue line in Fig. 3A is a linear fit for the low-power part of the data (where Pauli blocking is present) and returns a slope of 1.7 ± 0.17 [1/W].

The larger cloud size in Fig. 3B and the use of a small probe beam effectively reduce the average light intensity by 18% compared with the result shown in Fig. 3A [evaluated using a simple parameter-free model that accounts for thermal expansion (*13*)]. After correcting for this, the ratio of the slopes at low and high temperature is 0.63 ± 0.07 and agrees within its uncertainty to the Pauli suppression factor of 0.65 in Fig. 2. We have observed similar behavior for widely different parameters of the atom cloud and the probe beam. However, we find that this way of characterizing Pauli blocking is less direct (as the cloud changes

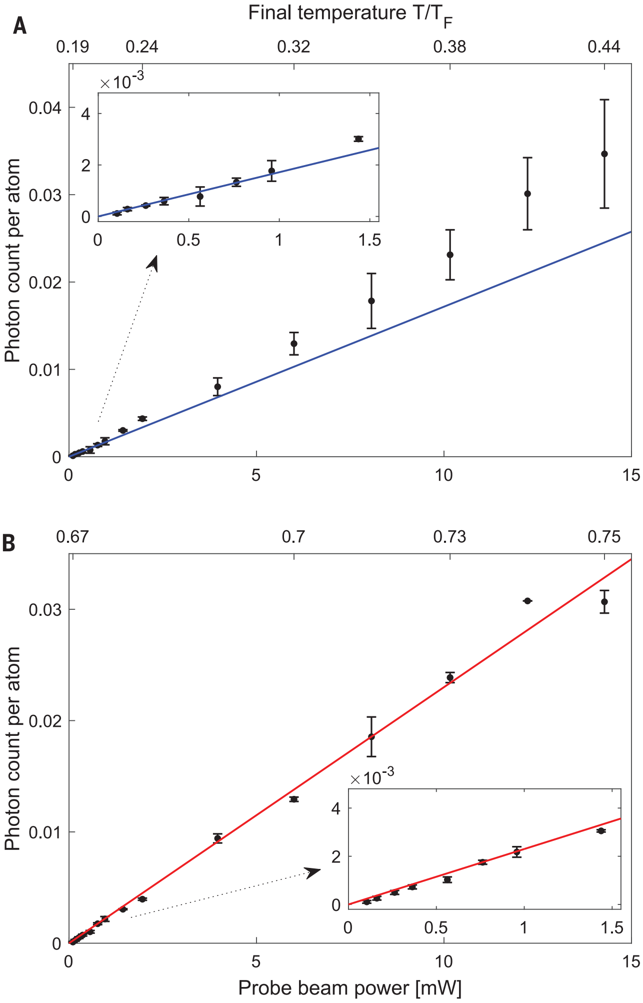


**Fig. 3. Light scattering as a function of probe beam power.** (**A** and **B**) We observe different slopes for low temperature, where Pauli blocking is present (A), and at high temperature (B). Solid lines are linear fits to the data [fit to the low-power part in (A) and to the full range in (B)]. The data in (B) were taken under similar experimental conditions as those in (A), but atoms were preheated to an initial temperature of $T/T_F$ = 0.67. At these temperatures, the blocking effect should be only ~5%. For the degenerate cloud, the signal is linear only for small laser power thanks to recoil heating. The top $x$ axis shows the final temperature ($T/T_F$), measured after releasing atoms from the trap. Detuning of the probe beam is $\Delta = -112$ GHz. Insets shows enlarged versions of the low-power region.

temperature during the probing) and more sensitive to fluctuations in the experiment than the method used for Fig. 2.

Because light scattering involves a (virtual) excited state, fermionic suppression of light scattering is related to Pauli suppression of spontaneous emission from an excited state embedded in a Fermi sea. The distinction between light scattering and spontaneous emission becomes important for an interacting system. It was shown theoretically that spontaneous emission in a zero-temperature Bose-Einstein condensate is enhanced by bosonic stimulation through the quantum depletion, whereas light scattering from a Bose-Einstein condensate is suppressed because the static structure factor $S(q) < 1$ owing to the phonon-dispersion relation (*17*).

So far, we have described Pauli blocking as a single-particle effect caused by Fermi statistics. However, because Fermi statistics create correlations between particles, one can also express Pauli blocking in terms of a pair correlation function. This will allow us to compare the suppression in our light-scattering experiment with other studies demonstrating fermionic suppression. The cross section of light (and also particle) scattering with momentum transfer $\hbar q$, $d\sigma/d\Omega$ is given by $S(q)$ times the single particle cross section $\sigma_0(q)$: $d\sigma/d\Omega = N\sigma_0(q)S(q)$ with $N$ representing the number of fermions. $S(q)$ is given by the Fourier transform of the density-density correlation function. A homogeneous system with $S(q) = 0$ would not scatter light. Uncorrelated classical particles show Poissonian fluctuations implying $S(q) = 1$. Suppression of light scattering off fermions is caused by suppressed density fluctuations, implying $S(q) < 1$. Suppression of density fluctuations in cold fermion clouds has been directly observed in previous studies (*18*–*20*), where the atomic density was shown to have sub-Poissonian fluctuations. This immediately implies reduced light scattering at small angles of order $k_F/k_i$. We have now extended this work by suppressing light scattering at all angles and directly detecting the scattered photons at a large angle. In the absence of longer-range correlations, the density-density correlation function is expressed by the pair correlation function $g(r)$, which is the normalized probability of detecting two particles at separation $r$, so the structure factor can be written as (*21*)

$$S(q) = 1 + n\int d^3r[g(r) - 1]e^{-iq\cdot r} \qquad (1)$$

For a nondegenerate quantum gas, $g(r) - 1 \neq 0$ for $r < \Lambda_t$, where $\Lambda_t$ is the thermal de Broglie wavelength (*22*)

$$g(r) \approx 1 \pm \exp\left(-2\pi r^2/\Lambda_t^2\right) \qquad (2)$$

resulting in

$$S(q) \approx 1 \pm D\exp\left(-q^2\Lambda_t^2/8\pi\right)/2^{3/2} \qquad (3)$$

where $D = n\Lambda_t^3$ is the peak phase space density reached around zero momentum. The averaged phase space density is $D/2^{3/2}$. The ± sign refers to bosons and fermions, respectively. The term "1" is the (normalized) contribution of the scattering of single atoms, whereas the second term is a consequence of nonvanishing interference terms involving light scattering by pairs of particles. Equation 3 can be generalized for degenerate gases (*22*) with the result that for fermions at zero temperature $S(q \to 0) = 0$, pair scattering completely cancels the scattering from single particles at angles for which $q \ll k_F$. This description emphasizes the central role of the pair correlations in the enhancement or suppression of light scattering. The pair correlation function has been directly observed in the suppression of elastic scattering (*23*) and inelastic scattering, including three-body recombination (*10*, *24*), which was crucial for the study of the BEC-BCS crossover (*25*), and the absence of interaction shifts in RF spectroscopy (*26*). We discuss in the supplementary materials (*13*) how the detuning dependence of inelastic light scattering can be used to map out the pair correlation function. Sometimes, these processes are described by Pauli suppression in the input channel (which is the p-wave), whereas the suppression of light scattering is regarded as Pauli suppression in the output channel. This distinction is correct, but it obscures the common origin of both effects, which are the pair correlations. Equation 1 is very general and applies also to colloidal particles with spatial correlations caused by interactions (*27*). It is only for noninteracting gases that quantum degeneracy is necessary to strongly modify the structure factor, and pair correlations have a one-to-one relation to bosonic enhancement or fermionic suppression.

We have directly observed Pauli blocking of light scattering. For our high densities, Pauli blocking is mainly limited by temperature, which can be lowered by an improved evaporation strategy addressing p-wave three-body recombination as the dominant loss mechanism (*10*). Pauli suppression can be used in quantum simulations to create fermionic samples that are less sensitive to heating when probed or manipulated by laser light. There are still many unresolved questions in how dense atomic samples scatter light, involving dipole-dipole interactions and superradiant scattering (*12*), and fermionic clouds with reduced incoherent scattering are a promising system for further studies.

## REFERENCES AND NOTES

1. K. Helmerson, M. Xiao, D. Pritchard, in *International Quantum Electronics Conference*, A. Owyoung, C. Shank, S. Chu, E. Ippen, Eds., vol. 8 of *OSA Technical Digest* (Optical Society of America, 1990).
2. J. Javanainen, J. Ruostekoski, *Phys. Rev. A* **52**, 3033–3046 (1995).
3. T. Busch, J. R. Anglin, J. I. Cirac, P. Zoller, *Europhys. Lett.* **44**, 1–6 (1998).
4. B. DeMarco, D. S. Jin, *Phys. Rev. A* **58**, R4267–R4270 (1998).
5. J. Ruostekoski, J. Javanainen, *Phys. Rev. Lett.* **82**, 4741–4744 (1999).
6. B. Shuve, J. H. Thywissen, *J. Phys. B At. Mol. Opt. Phys.* **43**, 015301 (2009).
7. B. O'Sullivan, T. Busch, *Phys. Rev. A* **79**, 033602 (2009).
8. R. M. Sandner, M. Müller, A. J. Daley, P. Zoller, *Phys. Rev. A* **84**, 043825 (2011).
9. I. Bloch, J. Dalibard, W. Zwerger, *Rev. Mod. Phys.* **80**, 885–964 (2008).
10. F. Ç. Top, Y. Margalit, W. Ketterle, *Phys. Rev. A* **104**, 043311 (2021).
11. J. D. Jackson, *Classical Electrodynamics* (Wiley, 1999).
12. B. Zhu, J. Cooper, J. Ye, A. M. Rey, *Phys. Rev. A* **94**, 023612 (2016).
13. Materials and methods are available as supplementary materials online.
14. K. B. Davis *et al*., *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
15. Z. Hadzibabic *et al*., *Phys. Rev. Lett.* **91**, 160401 (2003).
16. W. Ketterle, M. W. Zwierlein, in *Ultracold Fermi Gases, Proceedings of the International School of Physics "Enrico Fermi"*, M. Inguscio, W. Ketterle, C. Salomon, Eds. (IOS Press, 2008).
17. A. Görlitz, A. P. Chikkatur, W. Ketterle, *Phys. Rev. A* **63**, 041601 (2001).
18. C. Sanner *et al*., *Phys. Rev. Lett.* **105**, 040402 (2010).
19. T. Müller *et al*., *Phys. Rev. Lett.* **105**, 040401 (2010).
20. A. Omran *et al*., *Phys. Rev. Lett.* **115**, 263001 (2015).
21. D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford Univ. Press, 1987).
22. J. Bosse, K. N. Pathak, G. S. Singh, *Phys. Rev. E* **84**, 042101 (2011).
23. B. DeMarco, J. L. Bohn, J. P. Burke, M. Holland, D. S. Jin, *Phys. Rev. Lett.* **82**, 4208–4211 (1999).
24. D. S. Petrov, C. Salomon, G. V. Shlyapnikov, *Phys. Rev. Lett.* **93**, 090404 (2004).
25. T. Bourdel *et al*., *Phys. Rev. Lett.* **93**, 050401 (2004).
26. S. Gupta *et al*., *Science* **300**, 1723–1726 (2003).
27. A. Parola, R. Piazza, V. Degiorgio, *J. Chem. Phys.* **141**, 124902 (2014).
28. Y. Margalit, Figures data for "Pauli blocking of light scattering in degenerate fermions" (arXiv:2103.06921), *Zenodo* (2021); https://doi.org/10.5281/zenodo.5523336.
29. Y. Margalit, margalya/pauli-blocking: Release, version 1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.5523379.

## ACKNOWLEDGMENTS

We thank H. Son for comments on the manuscript. **Funding:** This work was supported by the NSF through the Center for Ultracold Atoms and grant no. 1506369, ARO-MURI Non-Equilibrium Many-Body Dynamics (grant no. W911NF-14-1-0003), AFOSR-MURI Quantum Phases of Matter (grant no. FA9550-14-10035), the ONR (grant no. N00014-17-1-2253), and a Vannevar-Bush Faculty Fellowship. **Author contributions:** All authors contributed to the concepts of the experiment; Y.M. and F.Ç.T. developed the sample preparation; Y.M. and Y.-K.L. designed the light-scattering setup and took and analyzed the data; and Y.M., Y.-K.L., and W.K. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data and code are deposited on Zenodo (*28*, *29*).

**SUPPLEMENTARY MATERIALS**

science.org/doi/10.1126/science.abi6153

Materials and Methods
Figs. S1 to S4
References (*30*–*33*)

19 March 2021; accepted 4 October 2021
10.1126/science.abi6153

---

QUANTUM GASES

# Pauli blocking of atom-light scattering

**Christian Sanner*†, Lindsay Sonderhouse†, Ross B. Hutson, Lingfeng Yan, William R. Milner, Jun Ye***

**Transition rates between coupled states in a quantum system depend on the density of available final states. The radiative decay of an excited atomic state has been suppressed by reducing the density of electromagnetic vacuum modes near the atomic transition. Likewise, reducing the density of available momentum modes of the atomic motion when it is embedded inside a Fermi sea will suppress spontaneous emission and photon scattering rates. Here we report the experimental demonstration of suppressed light scattering in a quantum degenerate Fermi gas. We systematically measured the dependence of the suppression factor on the temperature and Fermi energy of a strontium quantum gas and achieved suppression of scattering rates by up to a factor of 2 compared with a thermal gas.**

Radiative relaxation of an excited quantum system is a ubiquitous phenomenon: it makes fireflies glow, underlies the radiative recombination of electrons and holes in light-emitting diodes, and can be observed as gamma decay of nuclear isomers. The intimately related phenomenon of light scattering involves minimally populated excited states that often assume a virtual character (*1*, *2*). The blue sky is a direct manifestation of such a second-order excitation-emission process, with air molecules scattering sunlight. What all these light–matter interactions have in common is that the radiative decay depends on the final density of states for the joint emitter-photon system. For the photon channel in particular, it has been demonstrated (*3*–*5*) that manipulation of the density of vacuum modes through the use of an electromagnetic resonator modifies the emission and scattering of light. This so-called Purcell effect is now widely used in nanostructured devices (*6*). More than 30 years ago, it was suggested (*7*) that constraints imposed by the quantum statistics of an atomic medium could also modify spontaneous emission and light scattering. Fermi statistics requires the total wave function of a fermionic system to be antisymmetric, giving rise to the Pauli exclusion principle, which forbids indistinguishable fermions from occupying the same internal and external quantum states. Accordingly, if a sufficient number of ground-state fermionic atoms occupy all available external motional states into which an internally excited fermionic atom has to decay, this decay process will be blocked. Reducing light-scattering rates and extending the natural lifetime of an excited atomic state by embedding it inside a Fermi sea has been the theme of many theoretical studies and proposals (*8*–*17*), but experimental observation has been complicated by atomic properties and competing collective radiative behavior.

Spontaneous emission and light scattering are not synonymous. Even though the usual descriptions (*2*) of both processes involve the relaxation of an excited state (a populated atomic energy eigenstate and a virtual intermediate state, respectively), their properties and decay dynamics are different. In a noninteracting Fermi gas, however, the unavailability of final momentum states is the sole cause for single-particle modifications of radiative (re)emission properties and therefore universally affects all atom-light processes that transfer randomly directed recoil momentum to the atom (*13*).

Here we report a direct observation of Pauli suppression of light scattering using a quantum degenerate Fermi gas of strontium atoms. We confirm that the suppression becomes stronger as degeneracy is increased and when the Fermi energy approaches the photon recoil energy. By angularly resolving the photon scattering rate, we measured up to a factor-of-two reduction of light scattering in comparison to the value determined from a thermal ensemble. This distinctive manifestation of Fermi statistics connects the fundamental radiative property of atoms to their motional degrees of freedom subject to quantum statistics. The consequences of Pauli blocking of atomic motion have been previously demonstrated, including the suppression of collisions

---

JILA, National Institute of Standards and Technology and University of Colorado, Boulder, CO 80309, USA.
***Corresponding author. Email: christian.sanner@jila.colorado.edu (C.S.); ye@jila.colorado.edu (J.Y.)**
†These authors contributed equally to this work.

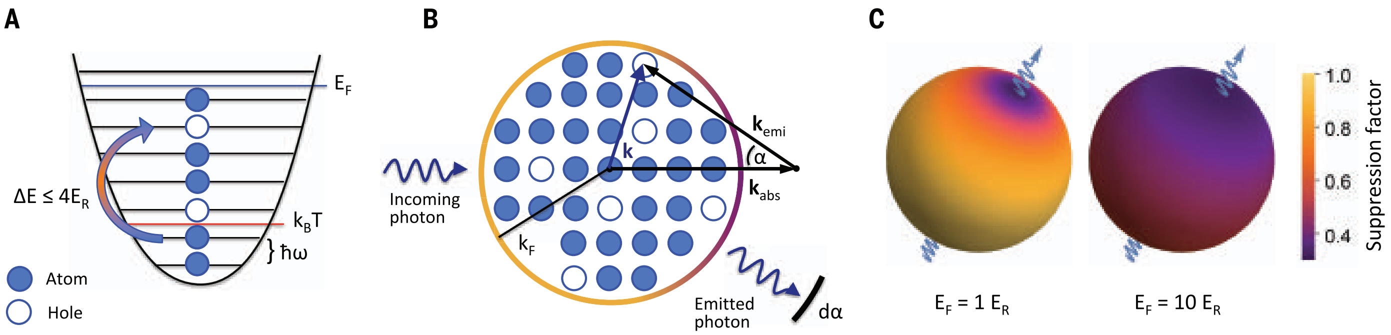


**Fig. 1. Light scattering on atoms embedded inside a Fermi sea.** (**A**) Indistinguishable fermions obey the Pauli exclusion principle. If the thermal energy $k_BT$ is sufficiently low, they fill almost all available harmonic oscillator states up to the Fermi energy $E_F$ with near-unity occupation. When scattering a photon, the atom's motional energy can change by up to four times the recoil energy $E_R$. (**B**) In momentum space, the atoms form a Fermi sea occupying most states up to the Fermi momentum $\hbar k_F$. Light scattering with reemission along $\alpha$ and total momentum transfer $\hbar\boldsymbol{k}$ can happen only if the final momentum state is not occupied by another ground-state fermion. Here, the atom initially in the center of the Fermi sea is assumed to scatter a photon. Atoms closer to the Fermi surface can reach a higher number of unoccupied momentum states. A detector covering a solid angle $d\alpha$ registers the emitted photon. (**C**) Pauli blocking leads to a characteristic angular distribution of scattered photons in the deeply degenerate regime (here, $T/T_F = 0.1$). For $E_F \sim E_R$ (left sphere), scattering is preferentially suppressed in a small cone around the forward direction, whereas $E_F \gg E_R$ (right sphere) causes strong suppression for all scattering angles $\alpha$. The suppression factor specifies the scattering rate relative to a non-Pauli-blocked sample.

(*18*), the direct observation of Fermi pressure (*19*), the onset of Hanbury Brown–Twiss anticorrelations (*20*, *21*), local antibunching (*22*–*24*), the suppression of chemical reactions between molecules (*25*), and the formation of Pauli crystals (*26*).

Figure 1 illustrates the key concept of the experiment and introduces the relevant energy scales (*27*). An ensemble of $N$ harmonically confined identical, noninteracting fermions of mass $m$ forms a quantum degenerate Fermi sea with near-unity occupation of the oscillator states if the thermal energy $k_BT$ is small compared with the Fermi energy $E_F = (6N)^{1/3}\hbar\omega$. The three-dimensional (3D) confinement is characterized by the mean trap frequency $\omega = (\omega_x\,\omega_y\,\omega_z)^{1/3}$, and $k_B$ and $\hbar$ denote the Boltzmann and reduced Planck constants, respectively. The Fermi temperature $T_F$ and Fermi wave vector $k_F$ are defined via $E_F = (\hbar k_F)^2/(2m) = k_BT_F$. If an atom at rest inside the Fermi sea absorbs a photon carrying momentum $\hbar\boldsymbol{k}_{abs}$, it gains a corresponding recoil energy $E_R = (\hbar\boldsymbol{k}_{abs})^2/2m$. Here, we consider the case of weak confinement with $\hbar\omega \ll E_R$ and treat Rayleigh scattering as a two-step momentum transfer process (*14*, *28*). Upon photon reemission, the atom experiences a randomly directed second momentum kick $\hbar\boldsymbol{k}_{emi}$, where $|\boldsymbol{k}_{emi}| = |\boldsymbol{k}_{abs}| = k_R$, resulting in a total momentum transfer $\hbar\boldsymbol{k} = \hbar\boldsymbol{k}_{abs} + \hbar\boldsymbol{k}_{emi}$. If, however, the corresponding motional state is already occupied by another atom within the Fermi sea, this decay channel is blocked, and light scattering will be suppressed. The relative temperature $T/T_F$ and the wave vector ratio $k/k_F$, where $\hbar k_F = (2mE_F)^{1/2}$ is the momentum space radius of the Fermi sea, determine the density of available final momentum states and hence the degree of blockade. Following Fermi's golden rule, one finds, using a local density approach (*11*, *14*), a relative scattering rate of

$$S(\boldsymbol{k}) = \frac{\int d^3\boldsymbol{p}\, d^3\boldsymbol{q}\, n_i(\boldsymbol{p},\boldsymbol{q})[1 - n_f(\boldsymbol{p},\boldsymbol{q})]}{\int d^3\boldsymbol{p}\, d^3\boldsymbol{q}\, n_i(\boldsymbol{p},\boldsymbol{q})}$$

Here the integrals cover the six-dimensional phase space spanned by three momentum dimensions $\boldsymbol{p}$ and three real-space dimensions $\boldsymbol{q}$. The initial and final state phase space cell occupations are given by $n_i = n_{FD}(\boldsymbol{p}, \boldsymbol{q})$ and $n_f = n_{FD}(\boldsymbol{p}+\hbar\boldsymbol{k}, \boldsymbol{q})$, where $n_{FD}$ is the Fermi-Dirac distribution for a harmonically trapped ideal Fermi gas, that is

$$n_{FD}(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{1+\xi^{-1}\exp\left[\left\{\sum_i p_i^2/(2m) + \sum_i m\omega_i^2 q_i^2/2\right\}/k_BT\right]}$$

The index $i$ runs over all three dimensions, and $\xi$ is the fugacity related to $T/T_F$ via $1/\mathrm{Li}_3(-\xi) = -6(T/T_F)^3$, where $\mathrm{Li}_3$ is the trilogarithm function. The expression for $S(\boldsymbol{k})$ counts all available final momentum states for a given momentum transfer $\hbar\boldsymbol{k}$ and averages over all initial states within the Fermi sea.

The above analysis suggests two pathways to observe pronounced Pauli suppression of light scattering. One can either prepare a Fermi gas with $E_F \gg E_R$ so that a substantial blockade is obtained for all emission directions, that is, for any momentum transfer up to the maximum $2\hbar k_R$ (homogeneously colored right sphere in Fig. 1C). Or one can relax this requirement and selectively observe only scattering events with a small momentum transfer so that $E_F \sim E_R$ is sufficient. The second approach, which we took in this study (gradient-colored left sphere in Fig. 1C), is straightforwardly realized in a small-angle light scattering configuration where a small number of atoms in the Fermi sea are optically excited, and a subset of the scattered photons is collected under a shallow angle $\alpha$ with respect to the excitation beam corresponding to a momentum transfer of $\hbar k = 2\hbar k_R \sin(\alpha/2)$.

A multitude of effects besides quantum statistics can influence radiation dynamics in a dense ensemble of emitters. Coherence, either externally imprinted or spontaneously established, can lead to super- and subradiant collective states that correspondingly exhibit super- and subnatural radiative lifetimes. Dicke superradiance (*29*, *30*), radiation trapping (*31*), multiple scattering (*32*), and other forms of coherent or incoherent collective scattering (*33*) all critically depend on the integrated optical density $\mathrm{OD} = \int \sigma n\, dl$ of the atomic gas. This attenuation parameter, derived from the single-atom scattering cross section $\sigma$ and density $n$, where length $l$ is measured along the direction of the incoming light, defines a parameter region where the gas is optically thin ($\mathrm{OD} \ll 1$) and single-particle scattering dominates over collective effects. Given that on resonance, $\sigma$ is on the order of the squared optical wavelength $\lambda^2 = (2\pi/k_R)^2$, and $k_F$ is tied to the 3D peak density via $n = k_F^3/6\pi^2$, it is not possible to satisfy the Pauli blocking criterion $k_F \sim k_R$ without violating the small OD requirement. Indeed, typical optical densities encountered in atomic Fermi gas experiments easily exceed 100. In this optically thick regime, multiple photon scattering strongly affects the light propagation inside the sample, as evidenced by the resonant fluorescence image displayed in Fig. 2B. Using off-resonant light reduces the effective scattering cross section and renders the atom cloud weakly absorbing for the incoming probe light at sufficient detuning. This, combined with differential observation strategies and avoiding light detection in the forward direction, minimizes the influence of collective scattering dynamics. Furthermore, unlike Pauli blocking, these competing effects are not dependent on quantum statistics and can therefore be observed using a thermal gas (*34*), instead of a quantum degenerate gas, to provide a baseline.

Our experiment started with the preparation of a $^{87}$Sr Fermi gas, as previously described (*35*). The $^1S_0$ $F = 9/2$ ground state is split into 10 magnetic spin states $m_F = -9/2, \ldots, 9/2$. Here, $F$ is the total angular momentum of the nuclear spin. This fully thermalized 10-component sample (*36*) contains 18,000 atoms per spin state confined in a crossed optical dipole trap with maximum radial trap frequencies of $\omega_x = \omega_y = 2\pi \times 120$ Hz and an axial confinement with $\omega_z = 2\pi \times 506$ Hz. This leads to a Fermi energy of $E_F = 440$ nK for each of the 10 Fermi seas; we reached temperatures down to 0.1 $T_F$. The small intercomponent s-wave scattering length $a = 97a_0$, where $a_0$ is the Bohr radius, makes the Fermi gas weakly interacting, with $k_F a = 0.06$, which is scaled by the number of spin states, increasing the total energy of the gas by <7% compared with a noninteracting gas. These weak repulsive interactions in principle affect $S(\boldsymbol{k})$, but their differential effect on temperature measurements and scattering rates is negligible (*36*) at the measurement precision achieved in this study. Under suitable conditions, any optical excitation returning to the ground state should experience a decay blockade. We performed specific measurements on the $^1S_0$-$^1P_1$ transition at 461 nm with a natural linewidth (*37*) of $\Gamma = 2\pi \times 30.4$ MHz and a recoil energy of $E_R = 520$ nK. Using a continuous weak-drive Rayleigh scattering scheme, the Fermi gas was exposed for 1 μs to a 1.2-GHz detuned drive beam that causes, on average, <10% of the atoms to undergo an excitation cycle. No evidence for inelastic scattering was observed.

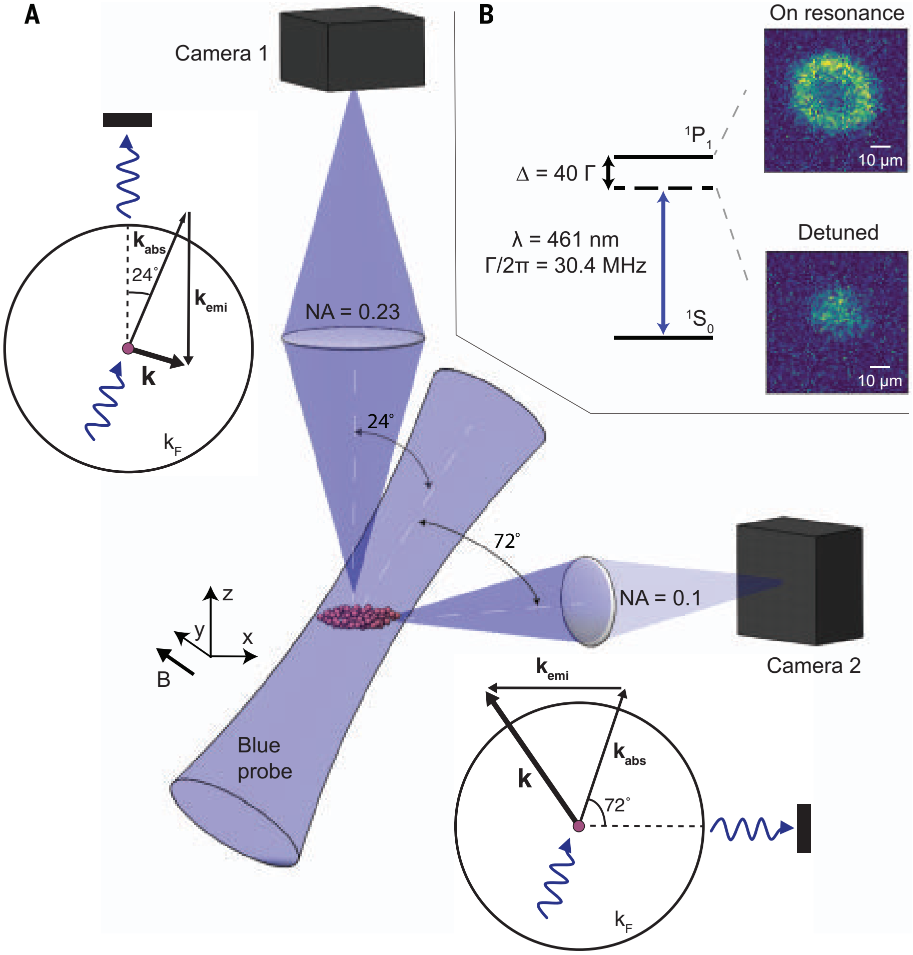


**Fig. 2. Experimental setup.** (**A**) Off-resonant probe light excites $^{87}$Sr atoms inside a Fermi sea. Scattered photons are collected simultaneously along two imaging axes under angles of 24° and 72°, with their numerical apertures (NA) shown. Small scattering angles correspond to a small momentum transfer with $k/k_F < 1$, whereas the transversal observation detects photons from scattering events with $k/k_F > 1$. The circularly polarized probe beam has an intensity of 5 $I_{sat}$, where the resonant saturation intensity is $I_{sat} = 41$ mW/cm$^2$. (**B**) On resonance, the atomic cloud is optically thick for the probe beam, and the image formed on camera 1 displays a hole in the cloud center owing to multiple scattering. At a detuning of $\Delta = 40\Gamma$, the atom cloud is optically thin, and the corresponding image resembles the atomic density distribution. The detuned frequency is used in the Pauli blocking experiment. A magnetic bias field of 3 G applied in the horizontal plane along the *y* direction defines the atomic quantization axis.

For the given drive beam and atom cloud parameters, the sample was optically thin with an effective OD of 0.02, corresponding to a forward transmission of 98%. As illustrated in Fig. 2, two detectors collected scattered photons simultaneously under off-axis angles of $\alpha_1 = 24°$ and $\alpha_2 = 72°$. These operational parameters were chosen to simplify the interpretation of our measurements. First, as discussed above, the detuning from resonance by 40Γ eliminates multiple scattering dynamics and furthermore avoids refractive lensing contributions (*38*) far off-axis. Second, the low excitation rate keeps the Fermi sea intact throughout the probe pulse exposure. Third, operating in the weak-drive limit ensures that there is no inelastic scattering contribution beyond the recoil-induced energy shift, i.e., Mollow triple-peaked fluorescence spectra and other strong-drive effects are negligible (*2*).

To systematically explore how quantum degeneracy affects light scattering, we started with a deeply degenerate gas of $T/T_F = 0.1$ and gradually heated it up to $T/T_F = 0.7$ through parametric confinement modulation while keeping the atom number and Fermi energy constant. Under these conditions, photons scattered off of the weak probe pulse were counted. To satisfy the requirement to minimally disturb the Fermi gas by the probe, the number of collected photons along the two detection axes was correspondingly low. Even for the axis with a high numerical aperture (NA) of 0.23, fewer than 200 photons were collected at full quantum efficiency. By operating the gated CCD detectors in a hardware binning mode that maps all detected photons onto a 3 by 3 superpixel array, we spatially and temporally isolated the signal from background contributions and maintained low readout noise. The measured

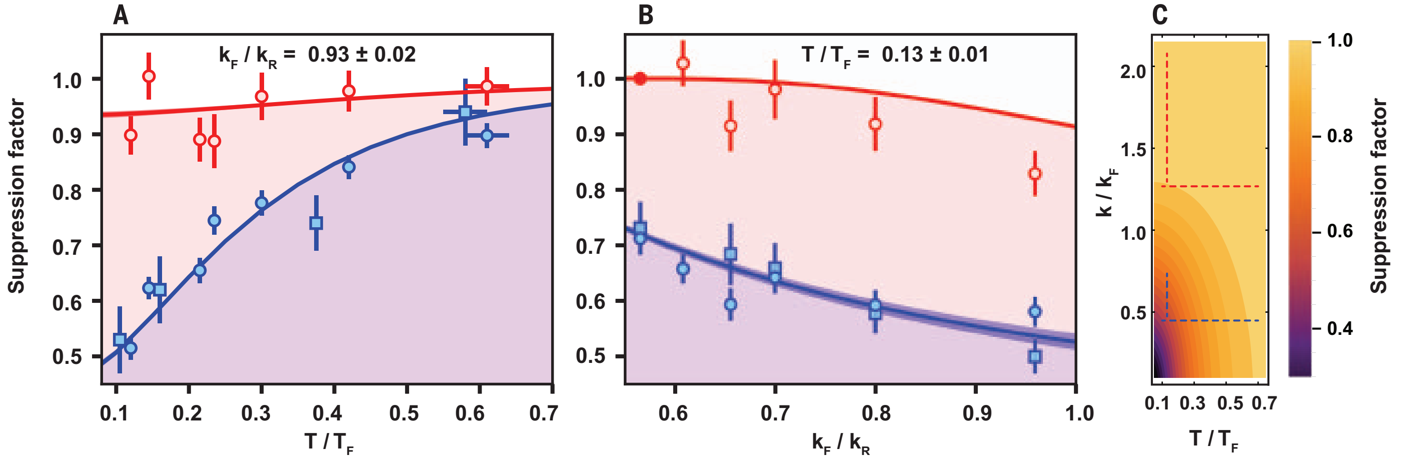


**Fig. 3. Suppression of light scattering in a $^{87}$Sr Fermi gas over a range of temperatures and Fermi momentums.** All measurements are performed with a 10-component Fermi gas containing 18,000 atoms per spin state. The scattering behavior is distinctly different for the two observation angles of 24° (blue circles and squares) and 72° (red circles). Raw photon counts are normalized with respect to measurements on non-Pauli-blocked reference samples (see text). Each circle data point is derived from 150 iterations of the experiment, and each square point results from 50 experimental runs. Solid theory curves are calculated with no free parameters (*14*). The widths of the theory lines reflect the experimental uncertainties of Fermi energy and temperature. The error bars are purely statistical and indicate one-standard-deviation confidence intervals. (**A**) At a constant Fermi wave vector of $k_F/k_R = 0.93$ ($E_F/E_R = 0.86$), the atom ensemble's scattering cross section decreases as the gas approaches deep quantum degeneracy with decreasing temperature. The suppression observed under 24° is pronounced, and the suppression factor reaches 50% at $T/T_F = 0.13$. In contrast, under 72°, the suppression is negligible. (**B**) At constant $T/T_F = 0.13$, $k_F$ is varied by adiabatically changing the confinement. A larger $k_F$ results in a stronger suppression. (**C**) The data reported in (A) and (B) are measured along four trajectories (dashed lines) through the parameter space spanned by $k/k_F$ and $T/T_F$. Depending on the scattering angle, $k$ varies between 0 and $2k_R$. Light collected under an off-axis angle of 24° corresponds to a momentum transfer $\hbar k < \hbar k_F$ for the given Fermi gas, leading to substantial reduction of the density of available final states. In contrast, for the 72° collection angle, the corresponding momentum transfer $\hbar k > \hbar k_F$. Thus most final states are not blocked, and scattering is not suppressed.

photon counts were compatible with ab initio estimates (*38*) within 30%, on the basis of the reported drive parameters and detection efficiencies. All relevant thermodynamic parameters were independently assessed through measurements on expanded gas clouds after time of flight.

The results of Pauli-suppressed scattering are shown in Fig. 3A. Under the shallow off-axis angle of 24°, where $k/k_F = 0.45$ for our Fermi sea, we find a strong dependence of the photon counts on $T/T_F$. To properly normalize the detector signal, that is, to convert the registered photon counts to a suppression ratio without introducing an arbitrary scaling factor, it was necessary to prepare an equivalent reference sample that was not Pauli-blocked. Because the weak optical confinement did not maintain a constant atom number for $T/T_F > 1$, we devised an alternative method to eliminate Pauli blocking: At 10 μs before applying the actual probe pulse, we exposed the Fermi gas to a 5-μs-long pre-pulse that destroyed the Fermi sea by randomly exciting atoms to momentum states beyond $k_F$. Consequently, light scattering was no longer Pauli-suppressed during the subsequent probe pulse. All round blue data points in Fig. 3, A and B, are normalized to a common reference photon count obtained from a single pre-pulse–exposed sample, as detailed in (*39*). Numerically integrating the expression for $S(\boldsymbol{k})$ at all probed temperatures reveals good agreement between experiment and theory. The simultaneously acquired measurements at an observation angle of 72° do not show a pronounced suppression and only exhibit a weak temperature dependence, as expected for $k/k_F = 1.27$, as most final momentum states lie outside of the Fermi sea.

To further verify that Pauli blocking is the mechanism responsible for the observed scattering behavior, we studied the dependence of the suppression factor on $k_F/k_R$ by varying the confinement while keeping the atom number and $T/T_F = 0.13$ constant, as displayed in Fig. 3B. Even at the shallowest confinement with $k_F/k_R = 0.57$ ($E_F/E_R = 0.32$), the momentum transfer along the 24° axis amounted to only $k/k_F = 0.74$, so that we still observed substantial suppression of light scattering. This is in contrast to the 72° case, where $k/k_F = 2.07$ at $k_F/k_R = 0.57$. If $k/k_F > 2$, Pauli blocking is negligible at any temperature, so we normalized all photon counts acquired under 72° (red data points in Fig. 3, A and B) with respect to this reference point (solid red circle in Fig. 3B).

The additional, blue square data points displayed in Fig. 3 were normalized by applying a pre-pulse separately for each data point and are in good agreement with the common-mode normalized measurements. In particular, with a 10-μs wait time before the probe pulse, the atomic density remains essentially the same with or without the pre-pulse. This ensures common-mode cancellation of collective scattering dynamics. In fig. S1, we present light scattering measurements with variable pre-pulse durations for a deeply degenerate ($T/T_F = 0.11$) and heated ($T/T_F = 0.58$) Fermi gas. These data confirm the scattering-induced destruction of the Fermi sea over the course of the pre-pulse application.

Spatially resolving the origin of the scattered photons within the Fermi gas, in addition to counting the total number of scattered photons along a given direction, provides a visual revelation of the Pauli blocking mechanism. For this purpose, we modified the 24° high-NA axis to deliver magnified images of the atom cloud and used a nonbinning CMOS camera with a low readout noise of two photoelectrons. The cylindrically symmetric atomic cloud at $T/T_F = 0.12$ with a diameter of ~20 μm was projected along the $z$ axis into a 2D image with 0.9-μm-wide pixels. Because a single pixel collects, on average, less than one photon, it is necessary to average hundreds of frames in order to derive a finely resolved scattering profile. Furthermore, we azimuthally averaged the mean image to obtain the radial profile (blue data points) shown in Fig. 4. Residual optomechanical drifts in the optical setup caused small displacements of the center of mass position of the cloud during the frame averaging period. This, together with the finite resolution of the imaging system, led to an effective $1/e^2$ pixel blurring on the order of 3 μm. To compare the observed scattering profile against theory predictions without introducing a free scaling parameter, we independently acquired in situ profiles of the Fermi gas through the same imaging setup via high-intensity fluorescent imaging (*40*). In this high-intensity limit, the light scattering rate is saturated and Pauli suppression effects are eliminated. Peak column density and width of these in situ profiles agree well with calculated atomic density distributions. Assuming no Pauli blockade, the scattering profiles at high and low probe intensities will therefore be related by a single rescaling factor that is given as the ratio of the total photon counts measured in the two cases. The so derived scattering profile expected without Pauli blockade for the weak probe beam is represented by the purple curve in Fig. 4. Momentum space integration plus one-dimensional line-of-sight integration of $S(\boldsymbol{k})$ yields a radially resolved suppression ratio for us to then determine the expected scattering profile with Pauli blockade (blue curve in Fig. 4). Except at the center of the cloud, where radial averaging does not substantially improve the signal-to-noise ratio, we find good agreement between calculated and measured profiles. Toward the outer rim of the cloud, the local Fermi energy drops so that light scattering is not suppressed anymore; the local suppression ratio will approach the thermal gas limit of 1. This happens, as seen with the blue theory curve, only in the outermost region, where the density is so low that the signal-to-noise level is insufficient to reliably determine a suppression ratio.

We have reported here a clear demonstration that Fermi statistics leads to strongly modified light scattering in a quantum degenerate system. The presence of a Fermi sea alters the final atomic motional mode spectrum and enables a direct observation of Pauli blockade of light scattering. Interpreted from a many-body system perspective, this experiment probes the structure factor (*41*) of a quantum degenerate Fermi gas. Defined as the Fourier transform of the spatial density–density correlation, the static structure factor $S(\boldsymbol{k})$ characterizes the linear response of a system to a perturbation with wave vector $\boldsymbol{k}$. Accordingly, the suppression of light scattering and the suppression of density fluctuations (*23*) in a Fermi gas are two interrelated manifestations of the same fundamental many-body physics.

Beyond the regime of Rayleigh scattering, we envision directly measuring a prolonged excited state lifetime using properly prepared quantum states (*42*). The capability of altering a fundamental decoherence mechanism will contribute to the quantum engineering of atom–light interfaces. In particular, custom-designed Fermi reservoirs can protect optical qubits at local nodes while facilitating cavity-free directional photon emission for efficient network connectivity. In the context of optical atomic clocks, this work could enable spectroscopic interrogation times exceeding the natural lifetime of the excited clock state and investigation of previously unexplored radiative properties of atoms.

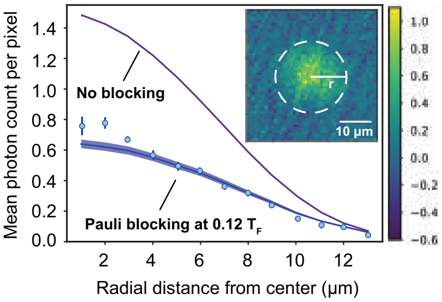


**Fig. 4. Spatially resolved light scattering from a trapped Fermi gas at $T/T_F$ = 0.12.** Azimuthally averaging the spatially resolved mean photon count (inset) from 1100 individual images obtained along the *z* axis yields a radial light scattering profile (blue data points). In situ column density images, separately obtained using a high-intensity fluorescent imaging technique, are used to predict the scattering signal for a nondegenerate gas (purple curve). The spatial profile of light scattering calculated for the $T/T_F$ = 0.12 ensemble (blue curve) agrees well with the measured data.

## REFERENCES AND NOTES

1. J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, 1967).
2. R. Loudon, *The Quantum Theory of Light* (Oxford Univ. Press, ed. 3, 2000).
3. E. M. Purcell, *Phys. Rev.* **69**, 681 (1946).
4. D. Kleppner, *Phys. Rev. Lett.* **47**, 233–236 (1981).
5. R. G. Hulet, E. S. Hilfer, D. Kleppner, *Phys. Rev. Lett.* **55**, 2137–2140 (1985).
6. L. Novotny, B. Hecht, *Principles of Nano-Optics* (Cambridge Univ. Press, 2006).
7. K. Helmerson, M. Xiao, D. E. Pritchard, in *International Quantum Electronics Conference*, A. Owyoung, C. Shank, S. Chu, E. Ippen, Eds., vol. 8 of OSA Technical Digest (Optical Society of America, 1990), paper QTHH4.
8. A. Imamoglu, L. You, *Phys. Rev. A* **50**, 2642–2645 (1994).
9. J. Javanainen, J. Ruostekoski, *Phys. Rev. A* **52**, 3033–3046 (1995).
10. T. Busch, J. R. Anglin, J. I. Cirac, P. Zoller, *Europhys. Lett.* **44**, 1–6 (1998).
11. B. DeMarco, D. S. Jin, *Phys. Rev. A* **58**, R4267–R4270 (1998).
12. J. Ruostekoski, J. Javanainen, *Phys. Rev. Lett.* **82**, 4741–4744 (1999).
13. A. Görlitz, A. P. Chikkatur, W. Ketterle, *Phys. Rev. A* **63**, 041601 (2001).
14. B. Shuve, J. H. Thywissen, *J. Phys. At. Mol. Opt. Phys.* **43**, 015301 (2009).
15. R. M. Sandner, M. Müller, A. J. Daley, P. Zoller, *Phys. Rev. A* **84**, 043825 (2011).
16. R. Onofrio, *Phys. Rev. A* **93**, 033414 (2016).
17. A. Piñeiro Orioli, A. M. Rey, *Phys. Rev. Lett.* **123**, 223601 (2019).
18. B. DeMarco, S. B. Papp, D. S. Jin, *Phys. Rev. Lett.* **86**, 5409–5412 (2001).
19. A. G. Truscott, K. E. Strecker, W. I. McAlexander, G. B. Partridge, R. G. Hulet, *Science* **291**, 2570–2572 (2001).
20. T. Rom *et al.*, *Nature* **444**, 733–736 (2006).
21. T. Jeltes *et al.*, *Nature* **445**, 402–405 (2007).
22. T. Müller *et al.*, *Phys. Rev. Lett.* **105**, 040401 (2010).
23. C. Sanner *et al.*, *Phys. Rev. Lett.* **105**, 040402 (2010).
24. A. Omran *et al.*, *Phys. Rev. Lett.* **115**, 263001 (2015).
25. L. De Marco *et al.*, *Science* **363**, 853–856 (2019).
26. M. Holten *et al.*, *Phys. Rev. Lett.* **126**, 020401 (2021).
27. S. Giorgini, L. P. Pitaevskii, S. Stringari, *Rev. Mod. Phys.* **80**, 1215–1274 (2008).
28. A. P. Vinogradov *et al.*, *Opt. Express* **29**, 2501–2520 (2021).
29. R. H. Dicke, *Phys. Rev.* **93**, 99–110 (1954).
30. M. Gross, S. Haroche, *Phys. Rep.* **93**, 301–396 (1982).
31. P. Weiss, M. O. Araújo, R. Kaiser, W. Guerin, *New J. Phys.* **20**, 063024 (2018).
32. M. C. W. van Rossum, T. M. Nieuwenhuizen, *Rev. Mod. Phys.* **71**, 313–371 (1999).
33. W. Guerin, M. T. Rouabah, R. Kaiser, *J. Mod. Opt.* **64**, 895–907 (2017).
34. S. L. Bromley *et al.*, *Nat. Commun.* **7**, 11039 (2016).
35. A. Goban *et al.*, *Nature* **563**, 369–373 (2018).
36. L. Sonderhouse *et al.*, *Nat. Phys.* **16**, 1216–1221 (2020).
37. A. Heinz *et al.*, *Phys. Rev. Lett.* **124**, 203201 (2020).
38. W. Ketterle, D. S. Durfee, D. M. Stamper-Kurn, in *Proceedings of the International School of Physics "Enrico Fermi,"* M. Inguscio, S. Stringari, C. E. Wieman, Eds. (IOS Press, 1999).
39. Materials and methods are available as supplementary materials.
40. M. T. DePue, S. Lukman Winoto, D. J. Han, D. S. Weiss, *Opt. Commun.* **180**, 73–79 (2000).
41. D. Pines, P. Nozieres, *The Theory of Quantum Liquids, Volume 1* (Addison-Wesley, 1988).
42. T. Bilitewski *et al.*, arXiv:2108.02819 [cond-mat.quant-gas] (2021).
43. C. Sanner, Replication Data for: Pauli blocking of atom-light scattering, version 3, Harvard Dataverse (2021); https://doi.org/10.7910/DVN/LHUKWV.

## ACKNOWLEDGMENTS

We thank P. Zoller, A. M. Rey, J. Thompson, and M. Holland for stimulating discussions and careful reading of the manuscript. **Funding:** Funding for this work was provided by the Defense Advanced Research Projects Agency, Air Force Office of Scientific Research, NSF Quantum Leap Challenge Institutes Office of Multidisciplinary Activities–2016244, NSF Phys-1734006, and National Institute of Standards and Technology. C.S. thanks the Humboldt Foundation for support. **Author contributions:** All authors contributed to carrying out the experiments, interpreting the results, and writing the manuscript. **Competing interests:** The authors declare that they have no competing financial interests. **Data and materials availability:** Data from the main text and supplementary materials are available through the Harvard Dataverse (*43*).

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abh3483
Supplementary Text
Fig. S1

3 March 2021; accepted 1 October 2021
10.1126/science.abh3483

METALLURGY

# Gradient cell–structured high-entropy alloy with exceptional strength and ductility

Qingsong Pan[1]†, Liangxue Zhang[1,2]†, Rui Feng[3]†, Qiuhong Lu[1], Ke An[3], Andrew Chihpin Chuang[4], Jonathan D. Poplawsky[5], Peter K. Liaw[6], Lei Lu[1]*

**Similar to conventional materials, most multicomponent high-entropy alloys (HEAs) lose ductility as they gain strength. In this study, we controllably introduced gradient nanoscaled dislocation cell structures in a stable single-phase HEA with face-centered cubic structure, thus resulting in enhanced strength without apparent loss of ductility. Upon application of strain, the sample-level structural gradient induces progressive formation of a high density of tiny stacking faults (SFs) and twins, nucleating from abundant low-angle dislocation cells. Furthermore, the SF-induced plasticity and the resultant refined structures, coupled with intensively accumulated dislocations, contribute to plasticity, increased strength, and work hardening. These findings offer a promising paradigm for tailoring properties with gradient dislocation cells at the nanoscale and advance our fundamental understanding of the intrinsic deformation behavior of HEAs.**

High-entropy alloys (HEAs), or alloys with multiple principal elements, have a near-infinite multicomponent phase space that can lead to unusual mechanical properties (*1*, *2*). Good strength and ductility, high work hardening, and exceptional damage tolerance have been achieved in some single-phase HEAs that have inherent concentration inhomogeneity created through tailoring of their chemical complexity (*3*–*6*). Moreover, engineering a spatial heterogeneous microstructure that consists of graded grain sizes, nanoclusters, multiphases, and so forth could also allow HEAs to achieve superior properties (*7*–*9*), similar to those that have been achieved in traditional heterostructured metallic materials (*10*–*12*). However, the long-lasting strength-ductility paradox for conventional metallic materials still exists for most HEAs (*7*, *8*, *13*).

The trade-off between strength and ductility for HEAs exists because the fundamental plastic-deformation features and mechanisms of HEAs reported so far are similar to those of conventional metals (*13*, *14*). Elementary line defects that carry plasticity—i.e., full dislocations and related interactions with different structural defects, such as high-angle grain boundaries (HAGBs) or twin boundaries (TBs)—are well understood in traditional metals (*15*–*18*). Notably, some unusual dislocation behaviors have been identified in HEAs with highly concentrated solid solutions, owing to local inhomogeneity with chemical short-range order (SRO) and spatially variable stacking fault energy (SFE) at the atomic scale (*14*). For example, the changing dislocation slip modes (*6*, *14*, *19*–*21*), as well as the enhanced friction resistance to motion or accumulation of dislocations (*5*, *22*, *23*), enabled by increasing local concentration fluctuations or local SRO at the nanoscale (generally <3 nm) (*5*, *6*, *24*), are believed to contribute to the improved mechanical properties.

We propose a heterogeneous gradient dislocation cell structure (GDS) in a stable single-phase face-centered–cubic (fcc) $Al_{0.1}CoCrFeNi$ HEA that contains randomly oriented, equiaxed fine grains (FGs) with an average diameter of ~46 μm. This alloy is a well-studied model material with a locally varied SFE of 6 to 21 $mJ/m^2$ (*25*). We found an unexpected and extremely high density of tiny stacking faults (SFs), twin nucleation, and accumulation-dominated plastic deformation in our GDS HEA upon initial application of tensile strain. This feature resulted in attractive strength and ductility properties compared with those of other HEAs (*10*, *26*, *27*).

To process HEA dog-bone–shaped bar specimens with a gauge diameter of 4.5 mm and a gauge length of 12 mm, we used a cyclic torsion (CT) treatment without surface tooling to form a sample-level multiscaled hierarchical dislocation structure under the imposed spatially gradient plastic strains from the surface to the core (*28*) (fig. S1). By controllably tuning CT parameters (i.e., torsion angle amplitudes of 20° and 6°), we prepared two different GDS samples with a constant torsion number of 200 cycles. We refer to these samples as GDS-H (Fig. 1; 20°) and GDS-L (fig. S2; 6°) on the basis of the imposed high (H) and low (L) accumulated torsion plastic strains (18.4 and 5.2, respectively, in the topmost layer; fig. S1E).

We focus on the GDS-H sample to illustrate the salient microstructural feature of a sample-level hierarchical dislocation structure. Grains from the surface to the core of the GDS-H sample are still homogeneously distributed, with characters of faceted grain morphologies, unchanged grain size, and random crystallographic orientations (Fig. 1, A and G), as evidenced by electron backscatter diffraction (EBSD) experiments with scanning electron microscopy (SEM). The same observations were made before CT (fig. S1C). The noticeably unchanged grain features after CT processing are in sharp contrast to those of traditional homogeneous or gradient nanostructures with severely refined grain sizes and higher densities of HAGBs that are produced by employing conventional severe plastic deformation strategies (*29*, *30*). Most grains in the sample core have the typical planar single-slip dislocation configuration in a relatively low dislocation density (fig. S3C), like that reported in the same HEA after tension strain (*25*). Well-developed single-slip–induced dislocation walls with low-angle boundaries (LABs) were occasionally observed in several grains (Fig. 1H and fig. S3D). Specifically, massive LABs with misorientations <15° were introduced in the topmost grain interior and spatially distributed such that they became smaller in volume fraction and larger in size with increasing the depth from the top surface (Fig. 1B; also schematically illustrated in Fig. 1C).

We randomly selected one grain separated by HAGBs in the topmost layer (~100 μm from the surface) of the GDS-H sample and used transmission electron microscopy (TEM) to investigate its dislocation structure. Figure 1D shows an example of the numerous equiaxed dislocation cells and cell walls with a small misorientation ranging from 0.7° to 4.8°, corresponding to massive LABs in Fig. 1B. We observed a misorientation gradient in the grain interior along the radial direction (white arrow in Fig. 1D), as a result of the imposed macro-level plastic-strain gradient upon the CT treatment (Fig. 1F) (*31*). We measured the cell diameter to be ~200 nm in the top surface, which gradually increases to ~450 nm for the dislocation wall structure in the core (fig. S3). Each cell wall (~40 nm thickness on average) is decorated with a high density of dislocations (~$1.7 \times 10^{15}$ $m^{-2}$) (*28*), but the cell interior has relatively fewer dislocations with curved morphologies (Fig. 1E and fig. S3). We achieved a similar dislocation structure in the GDS-L sample at a low cumulative plastic strain, exhibiting undeveloped dislocation cells with a larger average size of ~450 nm in the topmost surface (figs. S1 and S2). The dislocation cells in the GDS HEA sample are primarily caused by intensive multislip full dislocation interactions under a complex gradient stress-strain state (*32*) after a high cumulative torsion strain (fig. S1). We observed no visible SFs or deformation twins in the bulk GDS sample (Fig. 1 and figs. S2 and S3), indicating dislocation-controlled plastic deformation during the CT

---

[1]Shenyang National Laboratory for Materials Science, Institute of Metal Research, Chinese Academy of Sciences, Shenyang, 110016, P.R. China. [2]School of Materials Science and Engineering, University of Science and Technology of China, Shenyang, 110016, P.R. China. [3]Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [4]Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, USA. [5]Center for Nanophases Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [6]Department of Materials Science and Engineering, The University of Tennessee, Knoxville, TN 37996, USA.
***Corresponding author. Email: llu@imr.ac.cn**

†These authors contributed equally to this work.

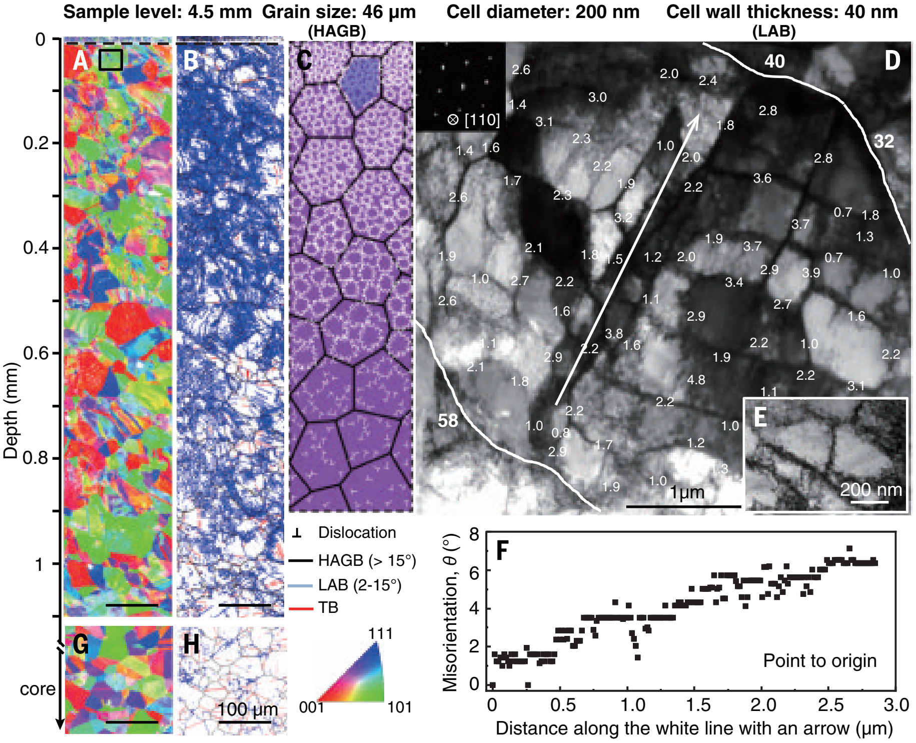


**Fig. 1. Typical microstructure and structural gradient of a gradient dislocation structure.** (**A** and **B**) Cross-sectional EBSD images of the GDS-H $Al_{0.1}CoCrFeNi$ HEA processed by cyclic torsion (CT) processing at a torsion angle amplitude of 20°, showing the distributions of a grain-scaled morphology, orientation (A), and three types of boundaries (HAGB, LAB, and TB) with different misorientation angles (B) within a depth of ~1.2 mm from the surface. (**G** and **H**) Same as (A) and (B) but for the core. (**C**) Schematic of GDS with a gradiently distributed low-angle dislocation structure. (**D**) Corresponding bright-field TEM image of dislocation structures at the topmost surface of the treated sample [indicated in (A) and (C)]. The misorientation angle of each cell wall, measured by means of an electron procession diffraction technique (*28*) in TEM, is indicated. The upper left inset is the corresponding SAED pattern. (**E**) A closer view of a typical dislocation cell structure. (**F**) Plots of misorientation-angle variation, measured with respect to the origin, across multiple cells at the topmost surface of the GDS-H HEA, along the direction of the white arrow in (D). Scale bars in (A), (B), and (G): 100 μm.

process, analogous to that observed in numerous HEAs with low SFEs (*13*, *25*, *33*, *34*).

Synchrotron x-ray diffraction (SXRD) scanning on the as-prepared GDS-H samples from the topmost surface to the core showed a spatial gradient–distributed dislocation density—up to $8.8 \times 10^{14}$ $m^{-2}$ in the topmost ~200 μm of depth (fig. S4). The GDS-H sample still exhibits a stable single-phase fcc structure, as evidenced by EBSD, TEM, and SXRD or neutron-diffraction results (Fig. 1 and figs. S5 and S8). Quantitative compositional analysis at the atomic scale by means of three-dimensional atom probe tomography (3D-APT) shows that the GDS structure is compositionally homogeneous without detectable elemental segregation at the cell wall (fig. S6).

The hierarchical dislocation cell structure results in substantially improved tensile properties (Fig. 2A). Tensile tests of both GDS samples show higher yield strengths ($\sigma_y$; 0.2% offset) of 362 ± 2 and 539 ± 26 MPa, about two to three times as strong as those of the FG and coarse-grained (CG) counterparts (185 ± 5 MPa and 138 ± 3 MPa, respectively) (*25*). In addition, we measured a high uniform elongation ($\delta_u$) up to 42.6 ± 0.2% for GDS-H, which is slightly reduced relative to that of FG counterparts without GDS (55.4 ± 3.4%). This behavior differs from the strength gain at the expense of ductility in most conventional metals and HEAs (*7*, *10*). Specifically, the GDS-H sample shows steady strain hardening with a slightly decreased work-hardening rate ($\Theta$)—from 1.28 GPa at a 3% strain to 0.99 GPa before necking—which is distinct from the trend of notably reduced $\Theta$ upon straining in its FG counterpart and the concave $\Theta$ shape of the CG counterpart (Fig. 2B) dominated by deformation twinning (*35*).

The GDS structure that we introduced leads to unexpected deformation-induced continuous hardening behavior that we found by measuring the microhardness [i.e., Vickers hardness ($H_V$)] along the depth from the surface to the cores of GDS and FG samples before and after application of tension (Fig. 2C). Owing to the presence of the sample-level gradient dislocation architecture in the whole cross-section, we identified a distinct, gradiently distributed $H_V$ from 3.7 GPa at the topmost surface to 2.2 GPa in the central region in GDS-H. These values are much higher than those of GDS-L (from 2.3 to 1.7 GPa). Compared to the uniform increase in $H_V$ for the FG sample after tensile straining, the hardening was particularly continuous in both tensioned GDS samples from the top surface to the core. This hardening feature is quite different from the deformation-induced continuous softening in conventional metals with gradient nanograins (*10*, *36*).

We attributed the pronounced increase in yield strengths to the nanoscaled dislocation cell unit with LABs (Fig. 1). Specifically, the contribution of LABs to strength has been demonstrated to be comparable with that of conventional HAGBs, and an ultrahigh hardness was observed in a nanolaminated Ni with LABs (*18*, *37*). For the topmost GDS surface layer of the GDS-H sample, the ultrahigh $H_V$ that we measured indicates that the massive low-angle dislocation cells are effective in resisting dislocation motion. This trend results from their structural features, including nanoscaled size and high density of dislocations. We argue that this feature, together with the obvious continuous hardening in the topmost GDS surface and the unusual work-hardening response of the whole GDS samples, indicates an enhanced-strengthening and ductilizing mechanism in gradient dislocation structures upon straining.

We further characterized the microstructural evolution at the top surface of GDS-H at strains of 3% (onset stage of steady work hardening) and 40% (later stage of plasticity) to unravel the intrinsic deformation mechanism of the gradient dislocation cell–structured HEA. At a strain of 3%, we found almost no detectable changes to the grain-level characteristics, including grain shape, size, and orientation (Fig. 3, A and H). However, the density of LABs was notably reduced in both the subsurface layer and the core. The presence of these dislocation patterns is attributable to a poorly developed or unstable state at relatively low cumulative plastic strains (*38*, *39*), whereas the pattern remains almost unchanged for the topmost grains (Fig. 3, B and I). Specifically, a widespread class of long, parallel lamellae bundles (average spacing of ~1.7 μm) was detected in the majority of the topmost grain, which decreases in number density as depth increases (Fig. 3, C and D, and fig. S7). The main components of these lamellae bundles are SFs, with a few TBs, as can be observed in selected-area electron diffraction (SAED) (Fig. 3E) and high-resolution TEM (Fig. 3F) images. We detected most parallel SFs and TBs (i.e., the dashed lines) across a dislocation wall (i.e., a white contrast region) (Fig. 3F). Close-up atomic-resolution TEM views show that

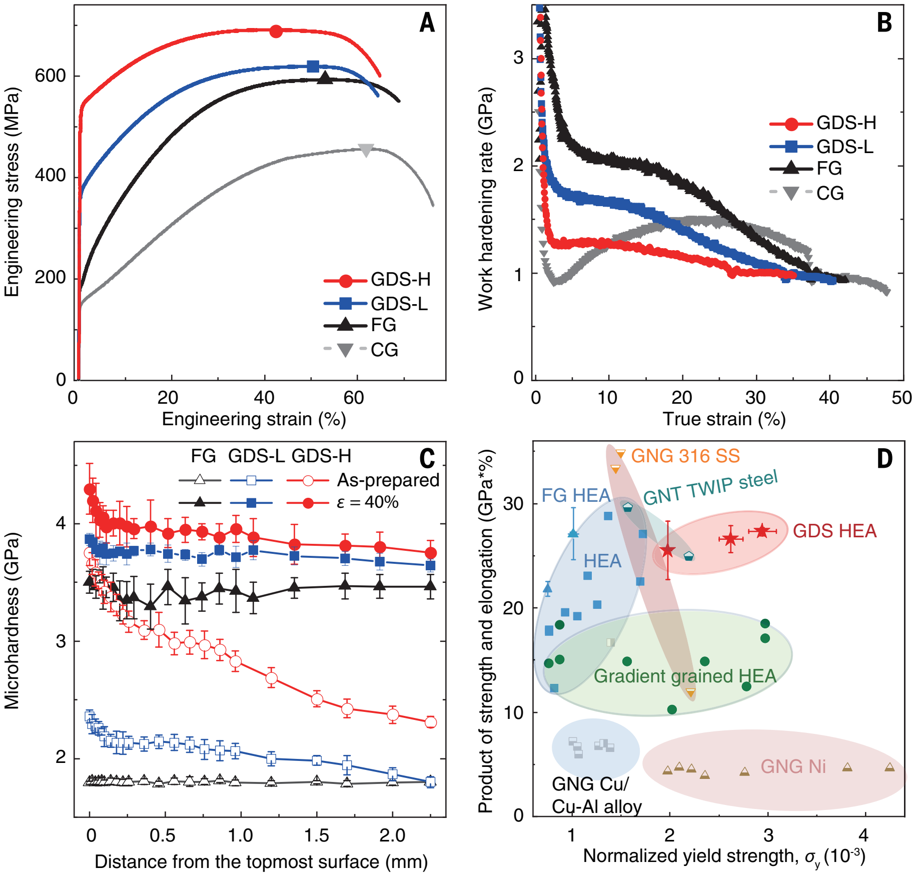


**Fig. 2. Mechanical properties of the GDS Al$_{0.1}$CoCrFeNi HEA.** (**A**) Tensile engineering stress-strain relations of GDS, CG, and FG samples. (**B**) Work-hardening rate and true strain relations of GDS samples compared with their homogeneous components. (**C**) Variations of measured microhardness along the distance from the top surface to the interior of GDS samples after CT processing. Measured values after application of a 40% tensile strain (ε) are also indicated. Error bars indicate SDs from 10 independent hardness measurements. (**D**) The product of strength and ductility versus yield strength normalized by the Young's modulus of the GDS Al$_{0.1}$CoCrFeNi HEA, compared with those of homogenous and gradient-grained structures and other metals and alloys with gradiently distributed nanograins and nanotwins reported in the literature (*10*, *36*, *52*–*55*). In (D), error bars represent SDs from more than three independent tensile tests. GNG and GNT denote the gradient nanograin and nanotwin, respectively; TWIP denotes twinning-induced plasticity; 316 SS denotes 316 stainless steel.

each individual long lamellae interface is essentially composed of numerous nanoscaled SFs or twin segments, ranging from several to tens of nanometers in length (Fig. 3G). Measurements of the thickness between adjacent SFs or TBs produce a much smaller value (2.9 nm on average), corresponding to a very high volume density (~$4.14 \times 10^8$ $m^2/m^3$) in the bundle. Moreover, the dislocation cells remain almost unchanged in shape and size. By contrast, the plastic deformation of most grains in the core of GDS-H is still dominated by intensive parallel full dislocations along {111} primarily slip planes (Fig. 3, J and K).

With the tensile strain increased to 40% before necking, the density of planar SF interfaces increases in volume fraction, prevailing in the grain interior (Fig. 4, A and B). Additional ultrafine nanotwins are also detected (Fig. 4C). The average spacing between adjacent SFs or TBs is small: ~4.4 nm (Fig. 4C). According to conventional wisdom, generation of both SFs and deformation twins is associated with the emission and slip of partial dislocations from GBs (*16*)—in this case, most likely from dislocation cell walls instead. Additionally, numerous short SFs along the other inclined {111} slip plane are widely detected in between neighboring SFs and TBs, thereby forming an unusual 3D SF-twin structure network (Fig. 4C).

We conducted in situ neutron-diffraction tensile experiments to confirm that the formation of SFs with detectable splitting of (111) and (222) planes happens at a very low tensile strain of ~5% for GDS-H samples, whereas it sets in at a high strain of ~30% for the FG counterpart (Fig. 4D). Our analysis of the neutron-diffraction results yields for the GDS-H sample a higher value of the SF probability than that of the FG counterpart upon straining (Fig. 4E), which indicates that SFs form more easily at an early deformation stage in GDS samples. Ex situ SXRD measurements further reveal that the enhanced probability of SFs and twins primarily stems from the contribution of GDS (fig. S8), consistent with our ex situ SEM and TEM observations (Figs. 3 and 4). By contrast, both SFs and deformation twins are rarely observed in FGs or at the core of GDS samples, even at a 40% strain, as confirmed by results in fig. S9 and in agreement with those reported in numerous single-phase fcc HEAs deformed at ambient temperature (*4*, *9*, *13*, *40*–*42*).

On the basis of the above results, we rationalize that the extremely high density of SFs and the TBs that mediate plastic deformation are primarily responsible for the superior mechanical properties of the gradient dislocation cell–structured HEA. As for the dominance of such dense SFs and TBs, which are not achievable in conventional metallic materials or most single-phase fcc HEAs in the early deformation stage at ambient temperature (*8*, *13*, *25*, *42*), we mainly attribute it to the chemical features of the HEA, nanoscaled dislocation cells with LABs, and their spatial gradient distribution.

Plastic deformation of conventional polycrystalline fcc metals is known to be mainly carried by the full dislocation slip and the intensive interactions within individual grains (*16*). As traditionally alternative ductilizing strategies of high-strength metals and alloys, twinning or phase transformation usually come into play, mostly through either tuning compositions to reduce the SFEs or lowering the deformation temperature (*16*, *43*). Under such circumstances, deformation twinning or phase transformation becomes prevalent, primarily owing to the ready partial dislocation glide on consecutive (or every other) {111} atomic planes in fcc metals (*16*, *43*). By contrast, for most single-phase fcc HEAs (*3*, *13*, *44*), neither twinning nor phase transformation dominates. Instead, the planar full dislocations usually dominate the plastic deformation, as detected in FGs (fig. S9) and the core of the GDS (Fig. 3K) after tension. Specifically, the SF dissociation distance of full dislocations in the deformed HEA is wider than in conventional metals—it can be as wide as 34 nm under tension tests at room temperature (*25*). We argue that the larger probability for partial dislocations and SF nucleation, rather than twinning or phase transformation, is closely related to the intrinsically spatially variable low SFE associated with the saliently rugged local atomic environments (*14*, *22*, *23*, *25*) and chemical SRO (*5*, *6*, *24*), inherent to HEAs, although we do not have direct evidence. The SF is a planar defect with energy relative to the lattice energy level where it forms. In particular, the increase in the lattice-distortion energy induced by high-density dislocations in the GDS can more readily adjust the local atomic positions to further decrease

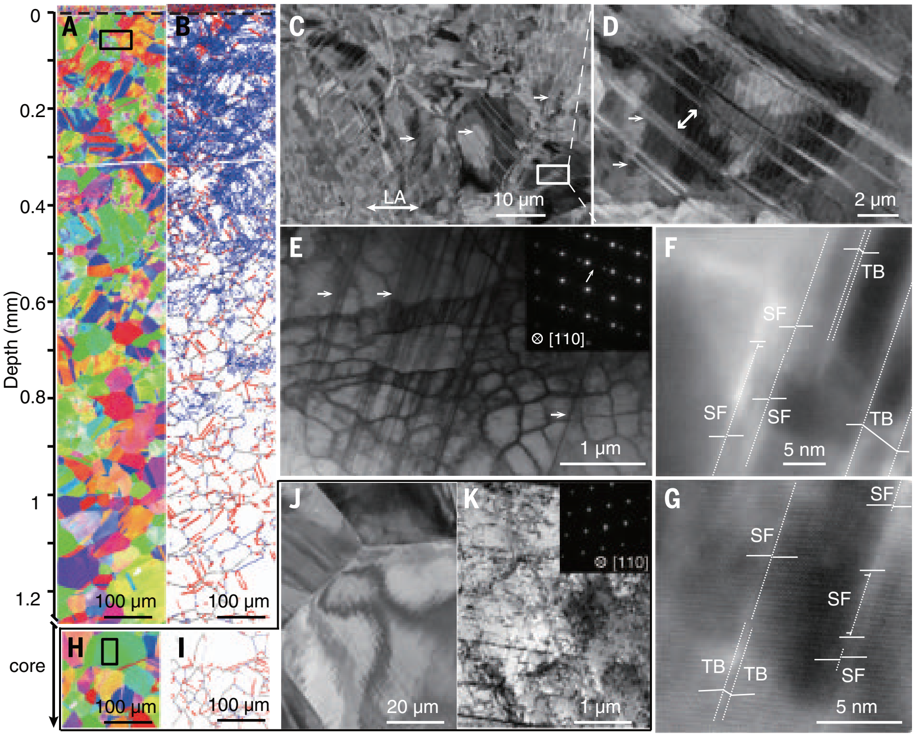


**Fig. 3. Deformation microstructure of the GDS-H $Al_{0.1}$CoCrFeNi HEA at a tensile strain of 3%.** (**A** and **B**) Cross-sectional EBSD images showing the distributions of the grain-scaled morphology, orientation (A), and three types of boundaries with different misorientation angles (B) within a depth of ~1.2 mm from the surface. (**H** and **I**) Same as (A) and (B) but for the core after tension. (**C** to **E**) Corresponding SEM [(C) and (D)] and bright-field TEM (E) images, which indicate the widespread occurrence of dense SF bundles (tens of micrometers in length), indicated by the white arrows, that cut through multiple dislocation cell structures. The inset in (E) is the corresponding SAED pattern that contains parallel streaks (along the [111] direction, denoted by the white arrow) from SFs. The two-headed arrow in (C) denotes the loading axis (LA). The two-headed arrow in (D) denotes the spacing of adjacent SF bundles. (**F**) Aberration-corrected high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) image taken from bundles in the vicinity of dislocation cell walls with a relatively low dislocation density, revealing an ultrahigh density of SFs and TBs. (**G**) Close-up HAADF-STEM image exhibiting numerous nanoscaled SFs or twin segments. The solid and dashed lines in (F) and (G) denote SFs or TBs and the (111) plane, respectively. (**J** and **K**) Corresponding SEM (J) and TEM (K) images at the sample core, indicating planar-slip–induced parallel dislocation morphologies.

the SFE (*45*), potentially increasing the SF width between partial dislocations.

In addition to the atomic-scaled compositional effect, the activity of the dislocation-dominated plastic deformation usually exhibits a strong (grain or cell) size dependence (*46*). As the characteristic size ($d$) decreases, the full dislocation activity is gradually inhibited, whereas the partial dislocation activity becomes more favorable. In terms of the classical Orowan relation (*15*, *47*), the critical stress required for initiating the full dislocation ($\tau_F$) and partial dislocation ($\tau_P$) as a function of cell size ($d$) can be expressed as

$$\tau_F(d) = \tau_0 + \frac{2\alpha\mu b}{d} \tag{1}$$

$$\tau_P(d) = \frac{\gamma}{b_p} + \frac{2\alpha\mu b_p}{d} \tag{2}$$

where $\tau_0$ and $\gamma$ are the lattice friction stress (57 MPa) and SFE (6 to 21 mJ/m$^2$) (*25*); $\alpha$ is a parameter that reflects the characteristic of dislocations (0.5 for an edge dislocation; 1.5 for a screw dislocation); $\mu$ is the shear modulus (80 GPa); and $b$ and $b_p$ are the Burgers vectors of full and partial dislocations (0.253 and 0.146 nm), respectively (*28*). With a maximum SFE of 21 mJ/m$^2$ and $\alpha = 1$ [roughly assuming that the dislocations in cell interiors have curved morphologies (Fig. 1D and fig. S2) belonging to a mixed type], the estimated critical size for the transition from full to partial dislocation activity is ~140 nm (fig. S10), which tends to become larger as the SFE decreases, evidently being comparable to the cell size in the topmost GDS layer (Fig. 1 and fig. S3). In essence, both the chemical and nanoscaled cell structural characters render partial dislocations highly favorable to full dislocations at ambient temperature (Fig. 3, F and G).

The sample-level gradient structure, with length scales spanning six orders of magnitude from the millimeter scale to the nanoscale [i.e., compositional fluctuation inhomogeneity or SRO, like those reported in the similar Al-Co-Cr-Fe-Ni systems (*5*, *48*)], plays a pivotal role in enabling activation of partial dislocations in our alloy. First, the built-in gradient distribution of dislocation cells causes progressive plastic deformation from the core to the topmost surface (fig. S11), effectively suppressing the overall strain localization. All nanoscaled dislocation cells jointly mediate plastic deformation, providing the spatial "playground" for activating partial dislocations and the resultant formation of SFs to undergo plastic deformation. Second, gradient plastic deformation in the multiscaled architecture is accompanied by a complex stress-strain state with the sample-level stress partition and the presence of the back stress, especially at the early stage of plastic deformation (*32*). By conducting tensile load-unload-reload tests, we determined the back stress of the GDS-H sample at a strain of 0.6% (close to 0.2% offset) to be as high as 260 MPa, which is ~50% of the yield strength and much higher than that of the FG counterpart, as expected. Consequently, such obvious enhanced strengthening originates from back-stress hardening associated with gradient plastic deformation of GDS, effectively contributing to a higher yield strength than that of the FG counterpart. Microscopically, the nanoscaled cells with a higher dislocation density at the walls tend to induce a stronger local stress field (than the measured sample-scaled macro-back stress), coupled with that caused by the local atomic distortions, thus endowing the internal driving force for partial dislocations slipping away from the equilibrium separation (*13*, *38*), with dense SFs stably left behind (Figs. 3, E to G, and 4).

Moreover, the ultrahigh density of dislocations at the cell boundaries naturally serves as an abundant, sustainable source for nucleating partial dislocations. Both large numbers and orientation-independence features of topmost grains that contain SFs (fig. S7) indicate that nucleation of SFs is not solely determined by the axial tensile stress state but is mostly controlled by complex multiaxial stress states in GDS upon straining.

Through engineering hierarchical nanoscaled GDS in a stable HEA, the deformation mechanism associated with extensive SFs and twins occurs even after a low tensile strain and is responsible for the steady work hardening and large tensile ductility. Such a high density of extremely fine SFs and TBs (~4.14 × 10$^8$ m$^2$/m$^3$) not only effectively mediates plastic strains, exhibiting SF-induced plasticity, but also substantially helps delocalize plastic deformation, coupled with gradient-induced strain delocalization. First, the penetrations of more and more dense SF and TB bundles across dislocation walls progressively subdivide stable

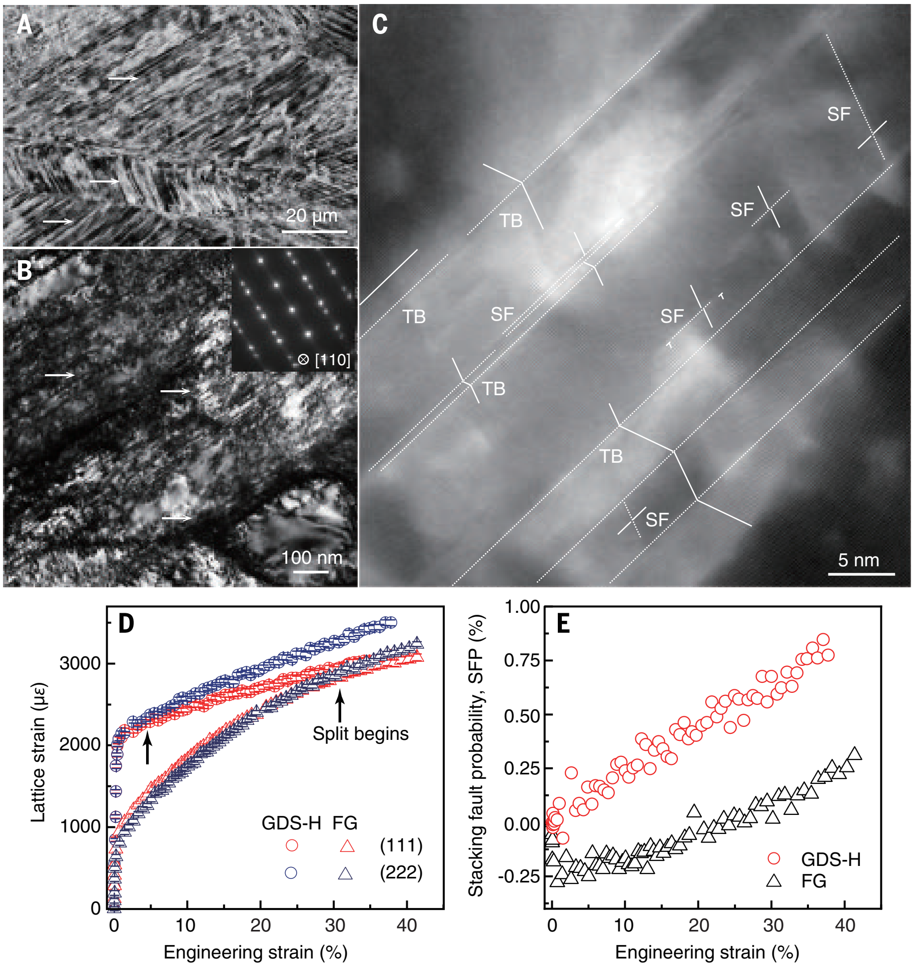


**Fig. 4. Deformation features at 40% tensile strain and in situ neutron-diffraction measurements during uniaxial tension of the GDS-H Al$_{0.1}$CoCrFeNi HEA.** (**A** and **B**) SEM (A) and TEM (B) images of a high density of SF bundles (indicated by white arrows) in the whole grain interior. The inset in (B) is the corresponding SAED pattern. (**C**) Typical atomic-resolution HAADF-STEM image showing an ultrahigh density of SFs and TBs with inclined short SFs. The solid and dashed lines in (C) denote SFs or TBs and the (111) plane, respectively. (**D**) Evolution of lattice strains for (111)//LA and (222)//LA grains in both GDS-H and FG samples versus engineering strain measured by in situ neutron-diffraction experiments. The error bars are obtained from the uncertainties of the single-peak fitting on *hkl* diffraction peaks (*hkl* denotes Miller indices). Black arrows indicate the occurrence of (111) and (222) splitting events. (**E**) Variation of the calculated SF probability (SFP) as a function of engineering strain in both GDS-H and FG samples.

dislocation structures and build in 3D bundle-cell networks (coupled with inherent, spatial local chemical fluctuations) that act as strong obstacles to dislocation slip (Fig. 3, C to G). In this case, intensive interactions among SFs or TBs and dislocations are promoted to induce considerable strengthening and work hardening, thus greatly counteracting the hastily reduced $\Theta$ trend of the GDS structure associated with rapid recovery of preexisting full dislocations and relatively limited capacity of dislocation multiplications at a low strain (Fig. 2B).

In addition, the progressively refined 3D SF-cell networks also act as sustainable sources for high-density full dislocation storage during uniaxial tension. The newly accumulated density of full dislocations in the GDS-H sample is substantially higher than that in the FG counterparts at the same tensile strain (fig. S4), still consequently offering a progressive and steady work-hardening response of GDS (Fig. 2B). Benefiting from the roles of such unexpected high-density linear and planar defects, bulk GDS exhibits better strain compatibility without strain localization and visible cracks in the grain interior or along GBs, thus resulting in high tensile plasticity comparable to that of the FG counterparts, at higher stress levels (Fig. 2A).

We compared yield strengths and the product of strength and ductility for GDS samples with the same compositional HEA (*40*, *41*, *49*–*51*) and other existing high-performance metals and alloys with gradient structures (*10*, *36*, *52*–*55*). The strength was normalized by the Young's modulus of the corresponding material. When plotted this way, we found that the GDS HEA achieves the best combination of uniform tensile elongation and strength. It is not surprising that conventional gradient nanograined metals and alloys with a high density of HAGBs suffer from structural coarsening, such as GB migration with concomitant coarsening (contributing to plasticity) and resultant softening, although these materials must sacrifice strength and stability to a certain extent (*10*). By contrast, monotonic strengthening with high, steady work hardening was achieved in the stable single-phase fcc GDS HEA with spatially gradient LABs, contributed by the aforementioned SF and twin behavior, with gradual structural refinement.

In summary, our observations show that engineering gradient LAB structures on a single-phase fcc Al$_{0.1}$CoCrFeNi HEA can help readily activate a mechanism to strengthen SF-induced plasticity, leading to exceptional strength and ductility. The discovery of such SF and twin behavior in the GDS HEA is essential for acquiring the common deformation features inherent to HEAs. It is also widely applicable to other HEA systems, particularly toward the achievement of better performance with superior properties, which are of fundamental and applied importance for advanced engineering applications such as automobiles, power stations, and aeronautic systems.

## REFERENCES AND NOTES

1. J. W. Yeh *et al*., *Adv. Eng. Mater.* **6**, 299–303 (2004).
2. B. Cantor, I. T. H. Chang, P. Knight, A. J. B. Vincent, *Mater. Sci. Eng. A* **375–377**, 213–218 (2004).
3. Y. Zhang *et al*., *Prog. Mater. Sci.* **61**, 1–93 (2014).
4. B. Gludovatz *et al*., *Science* **345**, 1153–1158 (2014).
5. Q. Ding *et al*., *Nature* **574**, 223–227 (2019).
6. R. Zhang *et al*., *Nature* **581**, 283–287 (2020).
7. E. Ma, X. Wu, *Nat. Commun.* **10**, 5623 (2019).
8. P. Sathiyamoorthi, H. S. Kim, *Prog. Mater. Sci.* 10.1016/j.pmatsci.2020.100709 (2020).
9. T. Yang *et al*., *Science* **362**, 933–937 (2018).
10. T. H. Fang, W. L. Li, N. R. Tao, K. Lu, *Science* **331**, 1587–1590 (2011).
11. X. Wu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14501–14505 (2015).
12. Y. M. Wang *et al*., *Nat. Mater.* **17**, 63–71 (2018).
13. W. Li *et al*., *Prog. Mater. Sci.* **118**, 100777 (2021).
14. E. Ma, *Scr. Mater.* **181**, 127–133 (2020).
15. J. P. Hirth, J. Lothe, *Theory of Dislocations* (Wiley, ed. 2, 1982).
16. M. A. Meyers, K. K. Chawla, *Mechanical Behavior of Materials* (Cambridge Univ. Press, ed. 2, 2009).
17. K. Lu, L. Lu, S. Suresh, *Science* **324**, 349–352 (2009).
18. K. Lu, *Nat. Rev. Mater.* **1**, 16019 (2016).
19. C. Lee, J. Brechtl, P. K. Liaw, *Metall. Mater. Trans. A* **52**, 2033–2093 (2021).
20. F. Wang *et al*., *Science* **370**, 95–101 (2020).
21. C. Lee *et al*., *Sci. Adv.* **6**, eaaz4748 (2020).
22. J. Ding, Q. Yu, M. Asta, R. O. Ritchie, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8919–8924 (2018).
23. Q. J. Li, H. Sheng, E. Ma, *Nat. Commun.* **10**, 3563 (2019).
24. X. Chen *et al*., *Nature* **592**, 712–716 (2021).
25. X. D. Xu *et al*., *Acta Mater.* **144**, 107–115 (2018).
26. K. S. Kumar, H. Van Swygenhoven, S. Suresh, *Acta Mater.* **51**, 5743–5774 (2003).
27. B. B. He *et al*., *Science* **357**, 1029–1032 (2017).
28. Materials and methods are available as supplementary materials.
29. M. A. Meyers, A. Mishra, D. J. Benson, *Prog. Mater. Sci.* **51**, 427–556 (2006).
30. W. Guo *et al*., *Acta Mater.* **170**, 176–186 (2019).

31. M. F. Ashby, *Philos. Mag.* **21**, 399–424 (1970).
32. Y. T. Zhu *et al.*, *Mater. Res. Lett.* **9**, 1–31 (2020).
33. Z. Li, K. G. Pradeep, Y. Deng, D. Raabe, C. C. Tasan, *Nature* **534**, 227–230 (2016).
34. F. Otto *et al.*, *Acta Mater.* **61**, 5743–5755 (2013).
35. S. W. Wu *et al.*, *Mater. Res. Lett.* **5**, 276–283 (2017).
36. Y. Lin, J. Pan, H. F. Zhou, H. J. Gao, Y. Li, *Acta Mater.* **153**, 279–289 (2018).
37. X. C. Liu, H. W. Zhang, K. Lu, *Science* **342**, 337–340 (2013).
38. H. Mughrabi, *Acta Metall.* **31**, 1367–1379 (1983).
39. P. Peralta, C. Laird, in *Physical Metallurgy*, D. E. Laughlin, K. Hono, Eds. (Elsevier, ed. 5, 2014), pp. 1765–1880.
40. M. Komarasamy, N. Kumar, Z. Tang, R. S. Mishra, P. K. Liaw, *Mater. Res. Lett.* **3**, 30–34 (2015).
41. D. Choudhuri, M. Komarasamy, V. Ageh, R. S. Mishra, *Mater. Chem. Phys.* **217**, 308–314 (2018).
42. M. Naeem *et al.*, *Sci. Adv.* **6**, eaax4002 (2020).
43. B. C. De Cooman, Y. Estrin, S. K. Kim, *Acta Mater.* **142**, 283–362 (2018).
44. E. P. George, D. Raabe, R. O. Ritchie, *Nat. Rev. Mater.* **4**, 515–534 (2019).
45. J.-W. Yeh, in *High-Entropy Alloys Fundamentals and Applications*, M. C. Gao, J.-W. Yeh, P. K. Liaw, Y. Zhang, Eds. (Springer, 2016), pp. 51–113.
46. R. J. Asaro, P. Krysl, B. Kad, *Philos. Mag. Lett.* **83**, 733–743 (2003).
47. E. Orowan, in *Symposium on Internal Stresses in Metals and Alloys* (Institute of Metals, 1948), p. 451.
48. L. J. Santodonato, P. K. Liaw, R. R. Unocic, H. Bei, J. R. Morris, *Nat. Commun.* **9**, 4520 (2018).
49. G. Chen *et al.*, *Scr. Mater.* **167**, 95–100 (2019).
50. N. Kumar *et al.*, *JOM* **67**, 1007–1013 (2015).
51. J. Yang *et al.*, *J. Alloys Compd.* **795**, 269–274 (2019).
52. H. W. Huang, Z. B. Wang, J. Lu, K. Lu, *Acta Mater.* **87**, 150–160 (2015).
53. J. Z. Long *et al.*, *Acta Mater.* **166**, 56–66 (2019).
54. H. T. Wang, N. R. Tao, K. Lu, *Scr. Mater.* **68**, 22–27 (2013).
55. J. J. Wang, N. R. Tao, K. Lu, *Acta Mater.* **180**, 231–242 (2019).

## ACKNOWLEDGMENTS

We thank S. Y. He for performing HAADF-STEM experiments, M. X. Yang for measuring the back stress, J. Xu for analyzing EBSD results, C. J. Zhang and L. F. Cheng for performing raw sample forging treatment, and J. Burns for performing APT experiments. **Funding:** Q.P. and L.L. acknowledge financial support from the National Science Foundation of China (NSFC; grants 51931010, 92163202, 52122104, and 52071321), the Key Research Program of Frontier Science and International partnership program (GJHZ2029). Youth Innovation Promotion Association (2019196), the Chinese Academy of Sciences (CAS), and LiaoNing Revitalization Talents Program (XLYC1802026). P.K.L. appreciates support from the National Science Foundation (DMR-1611180 and 1809640) and the US Army Research Office (W911NF-13-1-0438 and W911NF-19-2-0049). This study used resources at the Spallation Neutron Source (SNS), a US Department of Energy (DOE) Office of Science User Facility operated by the Oak Ridge National Laboratory (ORNL). Synchrotron diffraction was conducted at the Advanced Photon Source (APS), a US DOE Office of Science User Facility operated for the DOE Office of Science by the Argonne National Laboratory under contract DE-AC02-06CH11357. APT was conducted at ORNL's Center for Nanophase Materials Sciences (CNMS), which is a US DOE Office of Science User Facility. R.F. is grateful for support from Material Engineering Initiative (MEI) at SNS, ORNL. **Author contributions:** L.L. and P.K.L. initiated the project. L.L. supervised the project. Q.P. and L.Z. prepared the sample and performed the experimental tests. Q.L. conducted TEM observations. R.F., K.A., and A.C.C. performed the neutron and synchrotron diffraction experiments. R.F. carried out the diffraction line profile analysis. J.D.P. and R.F. performed the APT experiments and analysis. Q.P. and L.L. drafted the manuscript. All authors contributed to the discussions and revised the manuscript. **Competing interests:** A Chinese patent (grant number ZL201911044516.5) on the CT processing has been granted. **Data and materials availability:** All data generated or analyzed during this study are included in this article and its supplementary materials.

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj8114
Materials and Methods
Supplementary Text
Figs. S1 to S11
References (*56*–*83*)

3 June 2021; accepted 1 September 2021
Published online 23 September 2021
10.1126/science.abj8114

---

CORONAVIRUS

# COVID-19 transmission dynamics underlying epidemic waves in Kenya

**Samuel P. C. Brand[1,2,3]*†, John Ojal[1,4]*†, Rabia Aziza[2,3], Vincent Were[5], Emelda A. Okiro[6,7], Ivy K. Kombe[1], Caroline Mburu[1], Morris Ogero[1], Ambrose Agweyu[1,7], George M. Warimwe[1,7], James Nyagwange[1], Henry Karanja[1], John N. Gitonga[1], Daisy Mugo[1], Sophie Uyoga[1], Ifedayo M. O. Adetifa[1,8], J. Anthony G. Scott[1,8], Edward Otieno[1], Nickson Murunga[1], Mark Otiende[1], Lynette I. Ochola-Oyier[1], Charles N. Agoti[1], George Githinji[1], Kadondi Kasera[9], Patrick Amoth[9], Mercy Mwangangi[9], Rashid Aman[9], Wangari Ng'ang'a[10], Benjamin Tsofa[1], Philip Bejon[1,7], Matt. J. Keeling[2,3,11], D. James. Nokes[1,2,3], Edwine Barasa[5,7]**

**Policy decisions on COVID-19 interventions should be informed by a local, regional and national understanding of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission. Epidemic waves may result when restrictions are lifted or poorly adhered to, variants with new phenotypic properties successfully invade, or infection spreads to susceptible subpopulations. Three COVID-19 epidemic waves have been observed in Kenya. Using a mechanistic mathematical model, we explain the first two distinct waves by differences in contact rates in high and low social-economic groups, and the third wave by the introduction of higher-transmissibility variants. Reopening schools led to a minor increase in transmission between the second and third waves. Socioeconomic and urban–rural population structure are critical determinants of viral transmission in Kenya.**

After the first polymerase chain reaction (PCR)–confirmed case of COVID-19 in Kenya on 13 March 2020, the Kenyan government rapidly introduced measures aimed at suppressing severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission in the country. These measures included the closure of international borders, with the exception of cargo movement; closing of schools and other learning institutions; a ban on social gatherings and meetings; closure of places of worship, bars, and restaurants; a dawn-to-dusk curfew; mandatory wearing of masks in public places; physical distancing guidelines, including on public transportation; and restrictions on movement into or out of counties with high infection rates, including the two main Kenyan cities, Nairobi and Mombasa (*1*) (Fig. 1). Despite these measures, the rate of new COVID-19 cases grew in Kenya, indicating that measures had not been enough to consistently push the effective reproduction number $R(t)$ below 1. Moreover, serological surveillance indicated that a higher-than-expected fraction of the Kenyan population had been exposed to SARS-CoV-2 given the case reports at the time: June 2020 adjusted seroprevalences, based on blood donor samples from the Kenya National Blood Transfusion Services (KNBTS), were 5.6% for Kenya, 8% for Mombasa, and 7.3% for Nairobi (*2*).

Detected COVID-19 incidence in Kenya first peaked in early August 2020 during a period of relaxation of measures: the end of the Nairobi and Coastal counties (including Mombasa) lockdown (7 June 2020), and the resumption of international air travel (1 August 2020). A single-wave epidemic in Kenya peaking within 100 to 200 days after SARS-CoV-2 introduction into the country was initially predicted, based on assumptions that included a single population group and the development of immunity to reinfection (*3*–*6*). However, second and third waves occurred in mid-November 2020 and in March 2021, respectively. Multiple waves of COVID-19 incidence in high-income country settings have usually been associated with a relaxation of previous restrictions—for example, in the United Kingdom (*7*). More recently, the emergence of new variants has been associated with further waves of infection (*8*). In Kenya, and other countries in Africa, a temporal association between relaxation of restrictions and subsequent waves is implausible. Understanding the causation of such multiple waves is critical for forecasting hospitalization demand and the likely effectiveness of interventions, including vaccination strategy.

---

[1] Kenya Medical Research Institute (KEMRI)–Wellcome Trust Research Programme (KWTRP), Kilifi, Kenya. [2] The Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Warwick, UK. [3] School of Life Sciences, University of Warwick, Warwick, UK. [4] London School of Hygiene and Tropical Medicine (LSHTM), London, UK. [5] Health Economics Research Unit, KEMRI–Wellcome Trust Research Programme, Nairobi, Kenya. [6] Population Health Unit, Kenya Medical Research Institute–Wellcome Trust Research Programme, Nairobi, Kenya. [7] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [8] Department of Infectious Diseases Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. [9] Ministry of Health, Government of Kenya, Nairobi, Kenya. [10] Presidential Policy and Strategy Unit, The Presidency, Government of Kenya. [11] Mathematics Institute, University of Warwick, Warwick, UK.
***Corresponding author. Email: s.brand@warwick.ac.uk (S.P.C.B);jojal@kemri-wellcome.org (J.O.)**
†These authors contributed equally to this work.

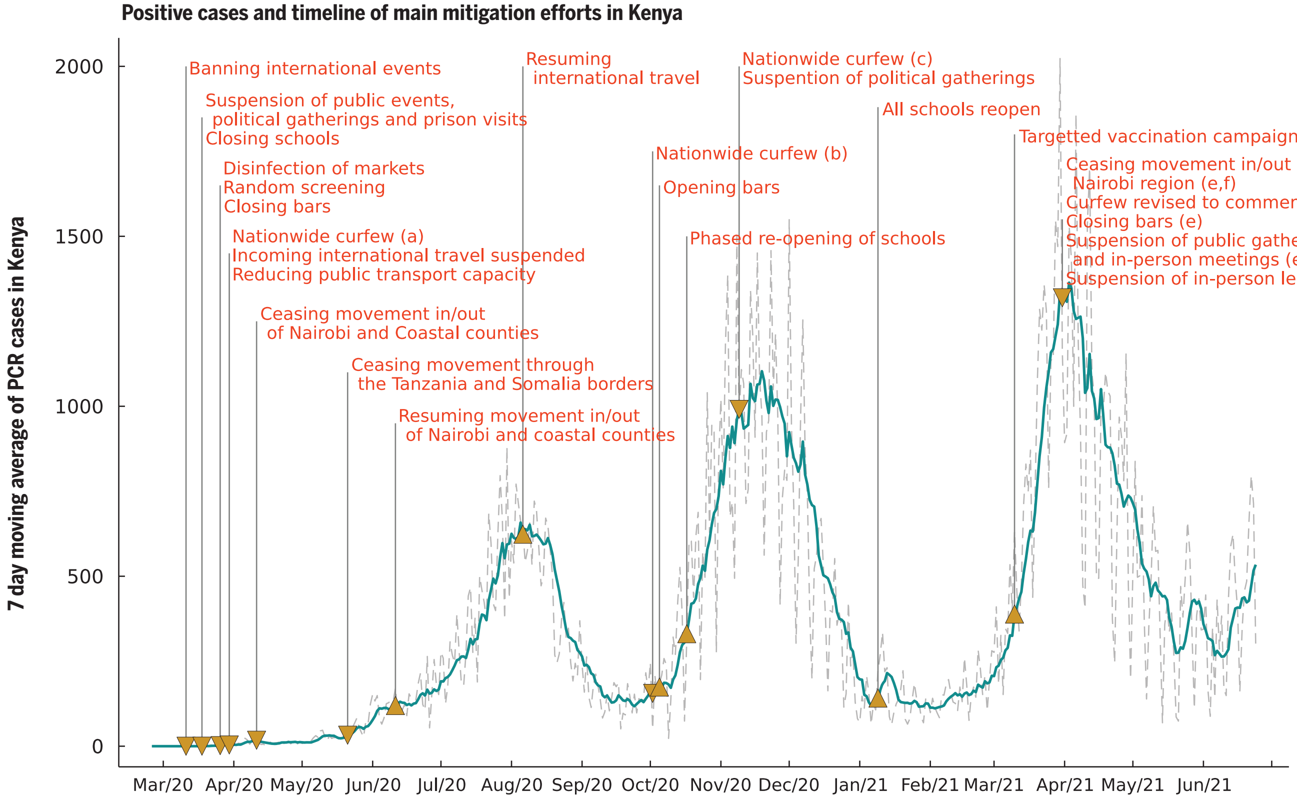


**Fig. 1. Seven-day moving average of daily positive PCR tests from the Kenyan national linelist and a timeline of the main mitigation events applied by the Kenyan government representing tightening (down-arrow) and relaxation (up-arrow) of measures.** (a) Curfew from 7 p.m. to 5 a.m.; (b) curfew from 11 p.m. to 4 a.m.; (c) curfew from 10 p.m. to 4 a.m.; (d) front-line workers and individuals older than 58 years (~1.2 million doses); (e) the region includes Nairobi, Kajiado, Machakos, Kiambu, and Nakuru; (f) this restricted movement into and out of the block of counties in (e) but not between these counties; (g) curfew from 8 p.m. to 4 a.m.

There are multiple potential explanations for sequential wave dynamics in COVID-19 incidence, which might be acting singly or in concert: social clustering (*9*), changing adherence to measures over time (*7*), seasonal effects on transmission (*10*), reopening of places of learning (*11*), lower transmission rates in rural settings leading to later peaks in those areas (*4*), and waning immunity after an infection episode, as well as the introduction of new SARS-CoV-2 variants that are more transmissible than previous strains and/or evade prior immunity acquired by natural infection (*12*). The decrease in cases following the peak of the first wave occurred at a time of relaxation of social distancing measures in Kenya (Fig. 1). Hence, the end of the first wave cannot be explained by the imposition of nonpharmaceutical interventions. In this work, we present evidence that the most plausible explanation for the pattern of cases and seroprevalence observed in Kenya is a combination of differential adherence to measures between subpopulations that we identify with lower and higher socioeconomic status (SES) in 2020 followed by a sharp increase in virus transmissibility in 2021, consistent with that observed in other countries affected by variants of concern, e.g., the United Kingdom (*13*) and India (*14*). Previous studies undertaken in sub-Saharan Africa at the level of individual country (*4*) or pan-African exploring the impact of climate (*15*) have not had the opportunity to integrate longitudinal PCR, serology, and Google mobility data.

We developed a county-specific, two-SES group, SEIRS-type transmission model, using a waning immunity rate derived from recent studies on the protectiveness of a natural infection to future reinfection (*16*–*19*). Our model includes, for each Kenyan county, (i) a SEIRS transmission model predicting new infections on each day, socioeconomic group, and county, which accounts for assortativity in infections—that is, the propensity for infected individuals to cause more intragroup infections compared to intergroup infections; and (ii) an observation model reflecting the data streams: PCR testing (positive and negative results), seroprevalence surveys, Google mobility data, and determined COVID-19 deaths. The model developed for this study differs from the standard SEIRS model with homogeneous mixing, adding the impact of new variants as detected by genomic surveillance and allowing the model to fit two socioeconomic groups in counties where this was supported by the data streams. We used a hierarchical approach to inferring the underlying epidemic trajectories in each of the 47 Kenyan semiautonomous counties by the following three steps: (a) grouping counties by similarity over a range of sociological and epidemiological metrics using machine learning; (b) for the 11 counties with a relatively high density of serology tests, we jointly inferred epidemiological model parameters for each county, e.g., (i) baseline $R_0$ for the county, (ii) the effect of schools being open on $R(t)$, (iii) the increase in transmissibility in February 2021 when B.1.1.7 lineage (Alpha variant) SARS-CoV-2 was first detected in Kenya (*20*), (iv) the fraction of the population in the higher SES group in each county and their assortative mixing rate, and (v) the fraction of cases reported for the county using Hamiltonian Markov chain Monte Carlo (*21*) with mildly informative priors; and (c) we inferred model parameters for the remaining 36 counties using informative priors for reporting fractions derived from a synthesis of the posterior distributions of counties grouped as similar to that county (see supplementary materials for details). We conducted formal model selection to compare one, two, and three SES group models, finding that the one-group model was an inadequate fit to the data, and the three-group model was not an improvement on the two-group model (see supplementary materials). We also conducted sensitivity analysis for different assumptions on waning immunity, finding consistent results for a range of scenarios (see supplementary materials).

The two-SES group transmission model was able to capture the timing and intensity of all three waves of Kenyan COVID case incidence and the trend of increasing proportion seropositive among Kenya National Blood Transfusion Service (KNBTS) donors (Fig. 2). We also validated the fitted model by comparing forecasts of seropositivity rates with those from data not used to infer model parameters. We used rounds 1 and 2 of the seropositivity survey using KNBTS donors for model parameter inference, collected during Mayto September 2020. Estimated seroprevalence among the Kenyan population, derived from the fitted two-SES group transmission model, was in good agreement with the out-of-sample round 3 of KNBTS seroprevalence data, collected January to March 2021 (Fig. 2). The Nairobi-specific epidemic trajectory inferred in this study agrees with seroprevalence estimates from a randomized survey from Nairobi and is congruent with the observation that it was public hospitals in Nairobi (favored by lower SES groups) that came under pressure in the first wave, whereas the second wave showed increased admission to private health facilities (figs. S7 and S8). As well as capturing the past trends of case reporting and seropositivity in Kenya, the fitted two-SES group transmission model accurately predicts the daily rate of new confirmed COVID-19 cases reported by the Kenyan Ministry of Health for the month after the censoring date of the PCR test data used to infer model parameters (Fig. 2).

The two-SES group transmission model reconciled the apparent paradox between evidence of the effectiveness of the rapidly introduced Kenyan measures in reducing mobility out of the home among Kenyan smartphone users, which was close to that observed in European and North American countries (fig. S1), and that case rates and fatality rates display two distinct waves in Kenya in 2020. The model provides an explanation for the end of the first wave through the depletion of susceptibles in geographically distinct, largely urban, subpopulations of lower SES. In some Kenyan counties (e.g., the urban counties Nairobi and Mombasa, and some of the semi-urban counties), we infer that a substantial group of people belong to the higher SES group whose mobility is well represented by Google smartphone data; a combination of school closures and reduction in mobility (by 44.5%; see supplementary materials) reduced the effective reproduction number sufficiently that newly infected people among the higher SES group were on average generating less than one secondary infection by April 2020 (Fig. 3). In the counties where the model finds evidence for distinct two-group dynamics (fig. S13), the model predicts low rates of intergroup infection transmission (posterior mean for the assortativity parameter estimates of 2 to 40 disassortative infections per 1000 potential infection events). We believe this can be ascribed to pandemic-induced changes in social behavior that enhanced intra-SES group infectious contacts (such as longer contact durations in families or local communities) and decreased inter-SES group infectious contacts as a result of, for example, avoiding public transport and cancelling domestic staff visits. The growth rate in cases, and relatively high levels of seroprevalence among KNBTS donors, are

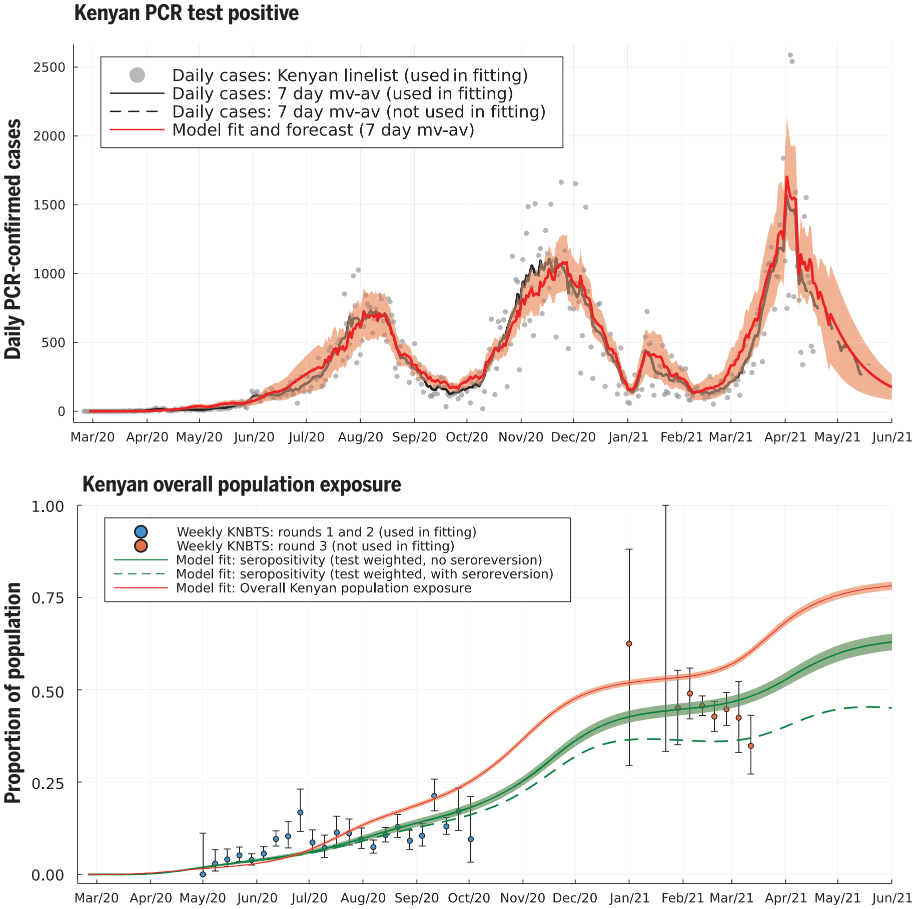


**Fig. 2. Daily PCR-confirmed COVID-19 cases (top) and weekly serology estimates from KNBTS donors with overall attack rate estimates (bottom).** Shown are daily numbers of PCR test positives from the Kenyan national linelist (top; gray dots are daily reports used in fitting the model, curves are 7-day moving averages). The model prediction for the 7-day moving average of daily case incidence (top; red dashed curve, shading shows 3-σ intervals) were derived from inference on the county-specific linelist PCR data and rounds 1 and 2 of the KNBTS serology survey (bottom; blue dots). Predictions before mid-April 2021 are back-calculations using known numbers of PCR tests per day, whereas after mid-April 2021, model predictions are forecasts that also estimate the number of PCR tests that will occur per day in each county. We show the next month of PCR test positive data, not used in fitting, as a validation of the model's short-term predictive accuracy (top; black dashed curve). Back-calculated model estimates of seropositivity (bottom; green solid curve) are shown with round 3 of the KNBTS serology survey data (bottom; red dots, not used in model inference). We also show back-calculated estimates of seropositivity under the assumption that median time to seroreversion (loss of detectable antibodies rather than loss of immunity) from infection was 1 year. Model estimates of overall Kenyan seropositivity are adjusted from county-specific estimates by weighting by number of serology tests in each county (over KNBTS rounds 1 and 2). The overall estimated Kenyan attack rate (population exposure) is shown as unweighted (bottom; red curve).

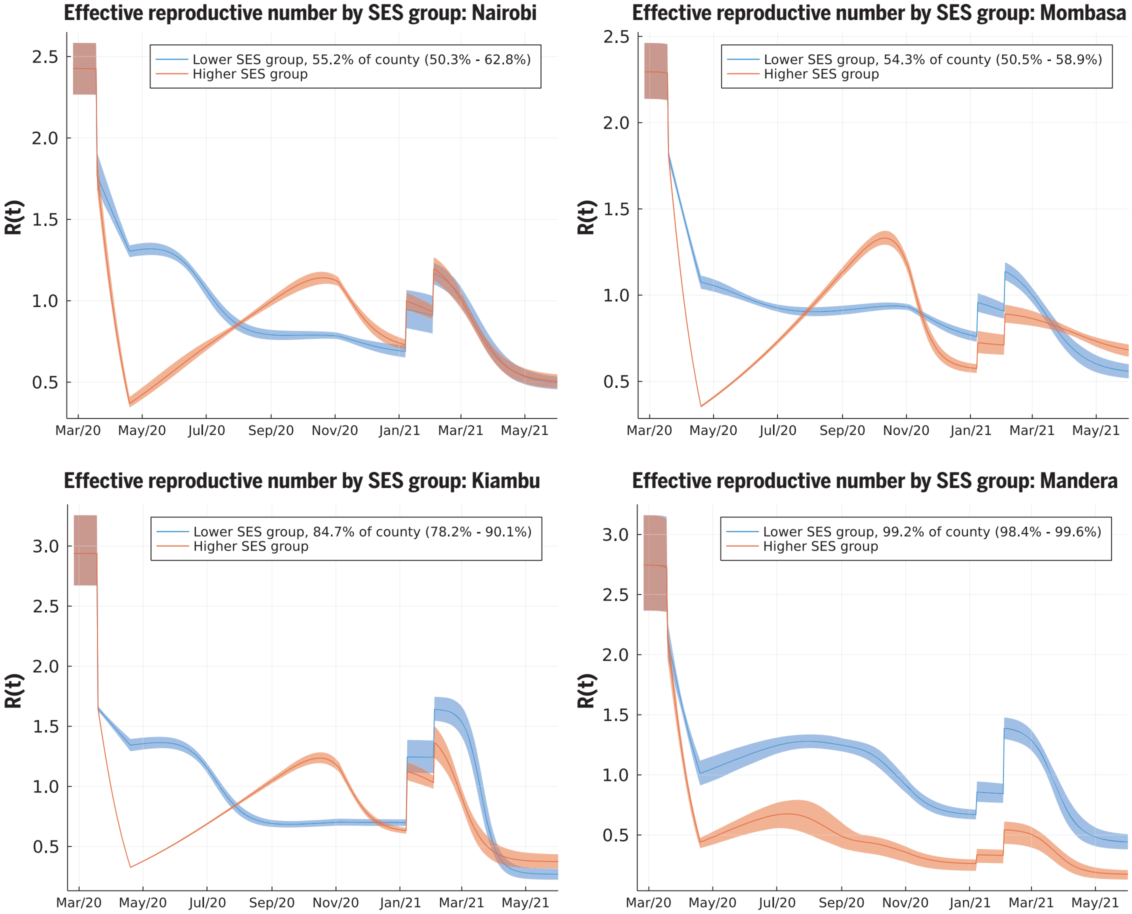


**Fig. 3. Effective reproduction number over time [*R*(*t*)] for lower and higher SES groups in four representative counties.** These include Nairobi (top left), Mombasa (top right), Kiambu (bottom left), and Mandera (bottom right). Nairobi and Mombasa are Kenya's two largest cities and form fully urban counties; Mandera county has a largely rural population and is remote from the main urban conurbations; Kiambu county borders Nairobi and has a ~60% urban population. The transmission model infers the proportion of the population in each SES group in each county. The highest proportion of higher SES group individuals are inferred to be in Nairobi and Mombasa out of all counties, whereas for Mandera county, very close to all individuals are inferred as being in the lower SES group, and the model effectively defaults to one group SEIRS transmission. The model inference for *R*(*t*) in Kiambu represents a county between these two extremes. In each county, the first discontinuous increase in *R*(*t*) is due to schools reopening, and the second is due to more transmissible variants becoming dominant in transmission.

explained by the rest of the population in the lower SES group having *R*(*t*) > 1 into May and June 2020 (Fig. 3). The model inferred that the reduction in mobility among the lower SES group was substantially less than among the higher SES group: The posterior mean estimate for reduction in mobility among the lower SES group in Nairobi was 13.8% [confidence interval (CI) 11.3–17.5%)] and in Kenya's second city Mombasa was 18.9% (CI 17.4–20.4%); posterior mean estimates for lower SES group mobility reduction across all 47 Kenyan counties had a median of 15.7% [interquartile range (IQR) 10.9–19.6%]. We assumed that school closures reduced *R*(*t*) for both SES groups equally. The inferred reduction in *R*(*t*) due to schools closing varied from county to county, and the median reduction in *R*(*t*) over counties was 13.5% (IQR 4.3–23.7%); the Nairobi estimate for school closure effect was 23.8% (CI 16.5–31.6%), and the Mombasa estimate for school closure effect was 20.2% (CI 15.2-25.2%) (Fig. 3).

The second wave in Kenya in 2020 was due to the superimposition of two trends: (i) in mainly urban areas, a second wave was triggered by the higher SES group returning to pre–COVID-19 mobility patterns by early November 2020 (fig. S1) and, therefore, *R*(*t*) was >1 for the higher SES group (Fig. 3, top and bottom left); and (ii) in more rural areas, the inferred size of the higher SES group was small, and *R*(*t*) was low but persistently >1 for the lower SES group [*R*(*t*) ~ 1 to 1.5] until November 2020 (Fig. 3, bottom right, and fig. S13; see supplementary materials). Low rates of mobility somewhat shielded the higher SES group from infection in the first wave among the lower SES group. Therefore, the lower SES group, in cities, suffered two waves in 2020, whereas the higher SES group effectively suffered one wave that peaked in late 2020 (Fig. 4). The overall detection rate was determined in part by the number of PCR tests performed each day, and this rate dropped in September 2020 (fig. S4). A consequence of the drop in the testing rate was that the case reporting shows a much sharper delineation between the first two waves (Fig. 2) than the underlying inferred infection rates (Fig. 4), which reveal only a moderate dip in infections in August–September 2020. By accounting for the delay between infection and COVID-19 fatality, and fitting SES group–specific infection-fatality-detection ratios (IFR detection; see materials and methods and supplementary materials) to each county, we found reasonable agreement between the predicted and observed timings of peak fatality rates in Kenya (Fig. 4). Overall, our model-based estimate was that only 11% of the total Kenyan population was in the higher SES group, whose mobility was well-described by Google mobility data, with the highest concentration of higher SES group individuals in the two main cities: 43.4% of the Nairobi population (CI 35.4–49.2%) and 40.3% of the Mombasa population (CI 35.0–45.4%). Additionally, we estimate that infections among the higher SES group were substantially more likely to be detected than among the lower SES group: The odds ratio for Nairobi for detection per case between higher and lower SES was 4.5 (CI 1.5–17.9), and for Mombasa for detection per case between higher and lower SES was 4.8 (CI 3.2–6.8). The

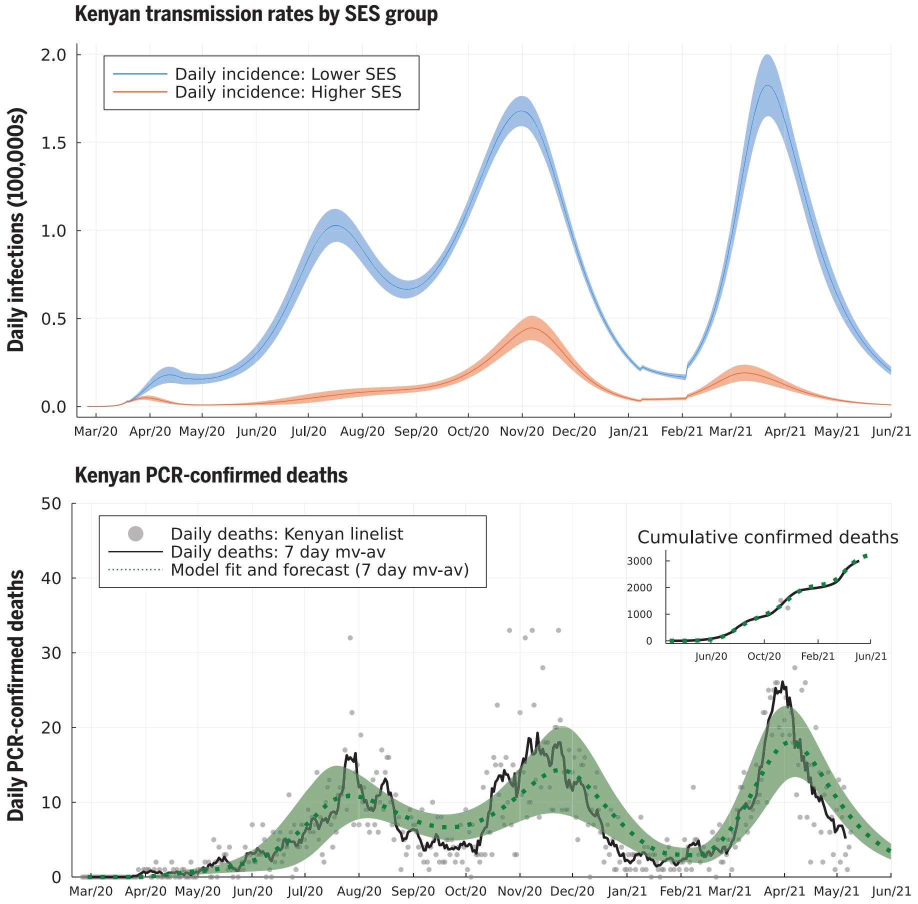


**Fig. 4. Model-inferred underlying true incidence rates by SES group relative to the whole Kenyan population size (top) and reported PCR-confirmed deaths due to COVID-19 disease (bottom).** The size of the upper SES group was estimated to be 11% of the Kenyan population, explaining the lower absolute rate of incidence (red curve) compared to the lower SES group (blue curve). We inferred that the lower SES group has experienced three waves of SARS-CoV-2 transmission, whereas the upper SES group has experienced two. The model fit for 7-day moving average (green dashed curve, with shading as 95% prediction intervals) is shown against the 7-day moving average for deaths reported in the Kenyan linelist (black curve). Cumulative observed and fitted deaths are shown in the top-right inset.

odds ratio between detection per infection in the two SES groups was inferred to be even more extreme across Kenya as a whole, with substantial variation from county to county: The median odds ratio estimate over counties was 18.5 (counties estimate IQR 2.5–27.9), although most counties had a small number of people in the higher SES group.

Fully reopening schools in January 2021 was associated with a slight increase in cases and deaths in Kenya, with a peak in January and early February 2021 (Figs. 3 and 4). However, school reopening does not explain the third wave in Kenya observed in March and April 2021, which was considerably larger than the increase in January and February 2021. The two-SES group model was not a sufficient explanation for a third wave, neither was loss of immunity or any detectable trend in mobility. The first cases of the more transmissible Alpha variant B.1.1.7 were identified in Kenya from mid-January 2021 (*20*). We investigated if the data supported an increase in transmissibility per infected person starting at the beginning of February 2021, as well as an influx of new exposed individuals representing domination of wild-type strains of SARS-CoV-2 by a fitter new variant. In the Kenyan urban counties, we found a credible range of increase in transmissibility of 15.0 to 46.6% [Nairobi 32.5% (CI 18.1–46.6%); Mombasa 22.8% (CI 15.0-31.2%)], and this was reflected in increased transmissibility estimates across Kenyan counties: The median over county estimates was 46.1% (IQR 31.6–72.9%). The fitted model predicted that this large increase in transmissibility will push the overall exposure to SARS-CoV-2 in Kenya from a back-calculated estimate of 53.5% in February 2021 to 78.1% by June 2021 (Fig. 2). The rate of seroreversion—that is, the loss of detectable antibodies rather than necessarily loss of protective immunity—has been identified as an important quantity for estimating population exposure prevalence from serological data (*22*). Because the serological data used for parameter inference was collected within 7 months of the first identified case in Kenya, we assumed that seroreversion was negligible over this period. However, assuming no future seroreversion led to closer agreement between model back-calculation and round 3 KNBTS data than assuming a median 1 year between infection and seroreversion (Fig. 2); that is, our modeling does not support the need to incorporate seroreversion to capture the true population exposure over the time scale of a year, unlike for Buss *et al.* (*22*). This finding highlights that seroreversion rate depends on the serological assay used (*23*) and cannot necessarily be extrapolated between settings. A negligible seroreversion rate may be more applicable for the enzyme-linked immunosorbent assay used in Kenya, where the cut-offs prioritize specificity over sensitivity (*2*, *24*).

Our modeling exercise provides a credible mechanistic interpretation of the three waves of COVID-19 in Kenya. We hypothesize the presence of two SES groups in each county and allow the model freedom to fit the relative proportion in each by county, inferred from locally collected PCR and serological test data. The model results support the notion of two SES groups in urban settings defined by highly assortative mixing (Nairobi, Mombasa, and predominantly counties near Nairobi), whereas for most rural counties, mixing was inferred to be less assortative and/or effectively all the population is in a single SES group (fig. S13). We invoke two key underlying assumptions. First, a stratified population differing in mobility (associated with lower and higher SES), and second, increased virus transmissibility compatible with competitive succession of a SARS-CoV-2 variant of concern in wave 3. A key simplifying assumption in this modeling study is that we assumed that the diversity of behaviors across the population in each Kenyan county can be reduced to stratifying into two groups with assortative mixing favoring transmission within group, and identifying these groups into lower and higher SES groups on the basis of similarity to mobility trends among smartphone users. We believe that this is a well-evidenced hypothesis for Kenya. Marked social and economic structuring has been described in Kenya; 36% of the population live below the national poverty line (*25*), and 55% live in informal settlements (*26*). Further, 83% of Kenya's labor market is informal, characterized by unstable and unpredictable daily wages (*27*). Lower socioeconomic groups have been identified as vulnerable to SARS-CoV-2 in the Global South because of residence in informal settlements at high population density, reduced access to sanitation, and dependence on informal employment requiring daily mobility (*28*, *29*). By contrast, the higher SES group with job security can work from home, socially distance, and readily access water and sanitation, thereby decreasing transmission. In Kenya, Google mobility data from smartphone users indicate a sharp decline in movement to settings outside of the home (fig. S1). We found that the two-SES group model used in this study was able to capture pattern of cases and seroprevalence in Kenya over the first three waves, despite radically simplifying the underlying social structure.

We predict the proportion of the Kenyan population exposed to SARS-CoV-2 to be greater than 75% by the beginning of June 2021 (Fig. 2), corresponding to around 39 million people. However, fewer than 4000 confirmed COVID-19 deaths and 180,000 confirmed SARS-CoV-2 infections have been identified as of 1 June 2021. We found that people among the lower SES group were likely to be even more undersampled than people among the upper SES group, as well as identifying wide regional variation in the detection rate. These results emphasize the necessity of community surveys of SARS-CoV-2 prevalence and exposure, and an investigation into the underreporting of mortality and severe disease due to COVID-19. Multiple waves have been observed in many other African countries that do not appear to be completely explained by the timing of restrictions, and because they also have in common similar socioeconomic groupings in urban centers, we speculate that the explanations found to be plausible in our model for Kenya may apply more widely.

The high population exposure suggests that a fourth COVID-19 wave in Kenya before the end of 2021 would only be likely if (i) a variant arises with substantial further enhancement in transmissibility or immune escape, such as the B.1.617.2 Delta variant (*30*), or (ii) there is substantial waning of immunity in those previously exposed. We predict that ~75% of the Kenyan population has been exposed to SARS-CoV-2 by June 2021. This will mitigate the death rate that might be expected in the future, but taking together (i) the markedly increased transmissibility of Delta variant, (ii) the potential for reinfection, and (iii) the experience of other countries despite prevalent vaccination, this scenario is entirely consistent with a fourth wave in Kenya. We conclude that our analysis, which triangulates PCR test, seroprevalence, and mobility and genomic data to develop a coherent explanation of the transmission dynamics of COVID-19, provides insight of public health importance in Kenya, including targeting health care capacity and pharmaceutical and nonpharmaceutical interventions.

## REFERENCES AND NOTES

1. E. Barasa *et al*., "Assessing the Indirect Health Effects of the COVID-19 Pandemic in Kenya" (2021); https://www.cgdev.org/publication/assessing-indirect-health-effects-covid-19-pandemic-kenya.
2. S. Uyoga *et al*., *Science* **371**, 79–82 (2021).
3. S. P. C. Brand *et al*., Forecasting the scale of the COVID-19 epidemic in Kenya. *medRxiv* 2020.04.09.20059865 [Preprint]. 16 April 2020.
4. B. Z. Diop, M. Ngom, C. Pougué Biyong, J. N. Pougué Biyong, *BMJ Glob. Health* **5**, e002699–e11 (2020).
5. S. Mwalili, M. Kimathi, V. N. Ojiambo, D. K. Gathungu, T. N. O.Achia, Age-structured Impact of Mitigation Strategies on COVID-19 Severity and Deaths in Kenya. *ResearchSquare* (2020); https://doi.org/10.21203/rs.3.rs-105797/v1.
6. J. Ojal, S. P. C. Brand, M. J. Keeling, D. James Nokes, *Wellcome Open Res.* **6**, 127 (2021).
7. M. J. Keeling *et al*., *PLOS Comput. Biol.* **17**, e1008619–e1008620 (2021).
8. H. Tegally *et al*., *Nature* **592**, 438–443 (2021).
9. B. F. Nielsen, L. Simonsen, K. Sneppen, *Phys. Rev. Lett.* **126**, 118301 (2021).
10. S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, M. Lipsitch, *Science* **368**, 860–868 (2020).
11. J. Panovska-Griffiths *et al*., *Lancet Child Adolesc. Health* **4**, 817–827 (2020).
12. T. Kirby, *Lancet Respir. Med.* **9**, e20–e21 (2021).
13. N. G. Davies *et al*., *Science* **372**, eabg3055 (2021).
14. J. Singh, S. A. Rahman, N. Z. Ehtesham, S. Hira, S. E. Hasnain, *Nat. Med.* **27**, 1131–1133 (2021).
15. R. E. Baker, W. Yang, G. A. Vecchi, C. J. E. Metcalf, B. T. Grenfell, *Science* **369**, 315–319 (2020).
16. R. A. Harvey *et al*., *JAMA Intern. Med.* **181**, 672–679 (2021).
17. V. J. Hall *et al*., *Lancet* **397**, 1459–1469 (2021).
18. J. Vitale *et al*., *JAMA Intern. Med.* **181**, 1407–1408 (2021).
19. F. Lumley *et al*., *Clin. Infect. Dis.* **73**, e699–e709 (2021).
20. E. Wilkinson, M. Giovanetti, H. Tegally, J. E. San, E. T. Cirulli, A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *medRxiv* 2021.05.12.21257080 [Preprint]. 13 May 2021.
21. A. Chatzilena, E. van Leeuwen, O. Ratmann, M. Baguelin, N. Demiris, *Epidemics* **29**, 100367 (2019).
22. L. F. Buss *et al*., *Science* **371**, 288–292 (2021).
23. The COVID-19 Community Research Partnership Study Group, Duration of SARS-CoV-2 Sero-Positivity in a Large Longitudinal Sero-Surveillance Cohort: The COVID-19 Community Research Partnership. *medRxiv* 2021.01.27.21250615 [Preprint]. 29 January 2021.
24. C. Bottomley *et al*., Improving SARS-CoV-2 cumulative incidence estimation through mixture modelling of antibody levels. *medRxiv* 2021.04.09.21254250 [Preprint]. 13 April 2021.
25. Kenyan National Bureau of Statistics Ministry Of Devolution National Planning, "KENYA - Kenya Integrated Household Budget Survey 2015-2016" (2018).
26. UN-Habitat, "UN-Habitat Support to Sustainable Urban Development in Kenya" (2017).
27. The World Bank Group, "Kenya Economic Update" (2016).
28. P. M. Macharia, N. K. Joseph, E. A. Okiro, A vulnerability index for COVID-19: spatial analysis to inform equitable response in Kenya. *medRxiv* 2020.05.27.20113803 [Preprint]. 28 May 2020.
29. J. Corburn *et al*., *J. Urban Health* **97**, 348–357 (2020).
30. I. Ferreira *et al*., SARS-CoV-2 B.1.617 emergence and sensitivity to vaccine-elicited antibodies. *bioRxiv* 2021.05.08.443253 [Preprint]. 9 May 2021.
31. S. Brand *et al*., SamuelBrand1/kenya-covid-three-waves: This is the version of the KenyaCoVSD module associated with the paper "COVID-19 Transmission Dynamics Underlying Epidemic Waves in Kenya," (Version KenyaCoVSDv1), Zenodo (2021); https://doi.org/10.5281/zenodo.5541433.

## ACKNOWLEDGMENTS

We thank the staff of the Kenya National Blood Transfusion Service and the blood donors, the county Rapid Response Teams and SARS-CoV-2 PCR testing laboratories, and data management staff of the Ministry of Health, Government of Kenya. This paper is published with permission of the Director of KEMRI. **Funding:** This work was supported by the UK Foreign, Commonwealth and Development Office (FCDO) and Wellcome Trust (220985/Z/20/Z); National Institute for Health Research (NIHR) Global Health Research Project GeMVi (17/63/82) and NIHR Global Health Research Unit on Mucosal Pathogens (16/136/46) (J.O.), using UK Aid from the UK Government to support global health research; Wellcome Trust Intermediate Fellowship awards (201866, 107568) (E.A.O. and L.I.O.-O.); MRC–DFID African Research Leader Fellowship (MR/S005293/1) (I.M.O.A. and C.M.); DFID–MRC–NIHR–Wellcome Trust Joint Global Health Trials Award (MR/R006083/1) (A.A.); Wellcome Trust Senior Research Fellowship (214320) and the NIHR Health Protection Research Unit in Immunisation (J.A.G.S.); Wellcome Trust (220991/Z/20/Z and 203077/Z/16/Z) (G.M.W.); and UKRI through the JUNIPER modeling consortium (grant no. MR/V038613/1). **Author contributions:** S.P.C.B.: Conceptualization, Formal Analysis, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; J.O.: Conceptualization, Formal Analysis, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; R.A.: Conceptualization, Formal Analysis, Methodology, Project Administration, Software, Writing – Review & Editing; V.W.: Conceptualization, Writing – Review & Editing; E.A.O.: Data Curation, Supervision, Writing – Review & Editing; I.K.K.: Formal Analysis, Methodology, Software, Writing – Review & Editing; C.M.: Formal Analysis, Methodology, Software, Writing – Review & Editing; M.O.: Data Curation, Writing – Review & Editing; A.A.: Conceptualization, Project Administration, Funding Acquisition, Writing – Review & Editing; G.M.W.: Conceptualization, Supervision, Funding Acquisition, Writing – Writing – Review & Editing; J.N., H.K., J.N.G., D.M.: Serology methodology; S.U.: Data Curation, Investigation, Writing – Review & Editing; I.M.O.A.: Conceptualization, Supervision, Writing – Review & Editing; J.A.G.S.: Conceptualization, Supervision, Funding Acquisition, Writing – Review & Editing; E.O., M.O., N.O.: Data Curation, Writing – Review & Editing; L.I.O.-O.: Data Curation, Investigation, Writing – Review & Editing; C.N.A., G.G.: Data Curation, Investigation, Writing – Review & Editing; K.K.: Project Administration, Writing – Review & Editing; P.A.: Project Administration, Writing – Review & Editing; M.M.: Project Administration, Writing – Review & Editing; R.A.: Project Administration, Writing – Review & Editing; W.N.: Project Administration, Writing – Review & Editing; B.T.: Project Administration, Writing – Review & Editing; P.B.: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; M.J.K.: Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Software, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; D.J.N.: Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; E.B.: Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing. **Competing interests:** The views expressed in this publication are those of the author(s) and not necessarily those of any of the funding agencies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. All authors declare no competing interests. **Data and materials availability:** All code and data for the transmission model underlying this study is accessible at the Github repository (https://github.com/SamuelBrand1/kenya-covid-three-waves and at Zenodo (*31*). All data are available in the main text or the supplementary materials. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

### SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk0414
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 and S2
References (*32–48*)
MDAR Reproducibility Checklist
Data S1 to S4

17 June 2021; accepted 4 October 2021
Published online 7 October 2021
10.1126/science.abk0414

CORONAVIRUS

# Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India

**Mahesh S. Dhar[1]†, Robin Marwal[1]†, Radhakrishnan VS[1]†, Kalaiarasan Ponnusamy[1]†, Bani Jolly[2,3]†, Rahul C. Bhoyar[2]†, Viren Sardana[2,3], Salwa Naushin[2,3], Mercy Rophina[2,3], Thomas A. Mellan[4], Swapnil Mishra[4], Charles Whittaker[4], Saman Fatihi[2,3], Meena Datta[1], Priyanka Singh[1], Uma Sharma[1], Rajat Ujjainiya[2,3], Nitin Bhatheja[2], Mohit Kumar Divakar[2,3], Manoj K. Singh[1], Mohamed Imran[2,3], Vigneshwar Senthivel[2,3], Ranjeet Maurya[2,3], Neha Jha[2], Priyanka Mehta[2], Vivekanand A[2,3], Pooja Sharma[2,3], Arvinden VR[2,3], Urmila Chaudhary[1], Namita Soni[1], Lipi Thukral[2,3], Seth Flaxman[5], Samir Bhatt[4,6], Rajesh Pandey[2,3], Debasis Dash[2,3], Mohammed Faruq[2,3], Hemlata Lall[1], Hema Gogia[1], Preeti Madan[1], Sanket Kulkarni[1], Himanshu Chauhan[1], Shantanu Sengupta[2,3], Sandhya Kabra[1], The Indian SARS-CoV-2 Genomics Consortium (INSACOG)‡, Ravindra K. Gupta[7,8], Sujeet K. Singh[1], Anurag Agrawal[2,3]*, Partha Rakshit[1]***

**Delhi, the national capital of India, experienced multiple severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreaks in 2020 and reached population seropositivity of >50% by 2021. During April 2021, the city became overwhelmed by COVID-19 cases and fatalities, as a new variant, B.1.617.2 (Delta), replaced B.1.1.7 (Alpha). A Bayesian model explains the growth advantage of Delta through a combination of increased transmissibility and reduced sensitivity to immune responses generated against earlier variants (median estimates: 1.5-fold greater transmissibility and 20% reduction in sensitivity). Seropositivity of an employee and family cohort increased from 42% to 87.5% between March and July 2021, with 27% reinfections, as judged by increased antibody concentration after a previous decline. The likely high transmissibility and partial evasion of immunity by the Delta variant contributed to an overwhelming surge in Delhi.**

After escaping relatively unscathed during the first wave of the COVID-19 pandemic, India witnessed a ferocious second COVID-19 wave starting in March 2021 that accounted for about half of global cases by the first week of May. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) had spread widely throughout India during the first wave, with the third national serosurvey in January 2021 finding that 21.4% of adults and 25.3% of 10- to 17-year-old adolescents were seropositive (*1*). Delhi, the national capital, was not included in the national serosurvey but had undergone multiple periods of high transmission in 2020 (Fig. 1A). In a district-wise stratified serosurvey conducted by the Delhi government in January 2021, overall seropositivity was reported to be 56.1% [95% confidence interval (CI), 55.5 to 56.8%], ranging from 49.1 to 62.2% across 11 districts (*2*). This level of seropositivity was expected to confer some protection against future outbreaks.

Despite high seropositivity, Delhi was among the most-affected cities during the second wave. The rise in new cases was exceptionally rapid in April 2021, increasing from ~2000 to 20,000 between 31 March and 16 April. This was accompanied by a rapid rise in hospitalizations and intensive care unit (ICU) admissions (Fig. 1B). In this emergency situation with saturated bed occupancy by 12 April, major private hospitals were declared by the state as COVID-care-only facilities, and senior medical students, including those from branches of alternative medicine, were pressed into service (*3*). Deaths rose proportionately (Fig. 1C), and the case fatality ratio (CFR), estimated as the scaling factor between time-advanced cases and deaths (Fig. 1D), was stable (mean, SD; 1.9, 0.3%). Population spread of SARS-CoV-2 is underestimated by test-positive cases alone (*1*, *2*). To better understand the degree of spread and the factors leading to the unexpectedly severe outbreak, we used all available data, including testing, sequencing, serosurveys, and serially followed cohorts.

In the absence of finely resolved or serial data from national and state surveys, we focused on data for Delhi participants of a national serosurvey of the Council of Scientific and Industrial Research (CSIR, India) employees and their family members (Fig. 2A and table S1). Samples were initially collected from the end of July to mid-September 2020 (phase I). Subsequently, second and third surveys were done in January and February (phase II) and the end of May to early July 2021 (phase III), bracketing the time period of interest. The cohort details and serosurvey methodology have been previously published (*4*).

Infection was determined by anti-nucleocapsid assay, which is not affected by immunization with spike protein–based vaccines. The presence of neutralizing antibodies to wild-type SARS-CoV-2 spike protein was estimated by a surrogate viral neutralization test (sVNT, Genscript). Previous results from the full cohort have been comparable to government serosurveys, but Delhi cohort values have been slightly lower. This may be due to an overrepresentation of members with the ability to reduce exposure and avoid public transport (Fig. 2A). Within these limitations, the Delhi cohort showed a rise in seropositivity from 14.7% in phase I (95% CI, 12.6 to 17.0%) to 42.1% in phase II (95% CI, 39.0 to 45.2%). About one-third of neutralizing antibody–positive subjects at phase I became neutralizing antibody–negative by phase II, with most showing declining inhibition on sVNT assays (Fig. 2B) (*5*). Phase III seropositivity increased to 87.5% (95% CI, 85.0 to 89.7%) among unvaccinated subjects. New infections between March and June 2021 are thus likely to have vastly exceeded known cases. Among 91 previously infected subjects with serial measurements at three time points, including phase III (T3), 25 (27.5%; 95% CI, 18.4 to 37.5%) had a pattern of declining antibody concentration between T1 and T2, followed by a sharp rise at T3, indicative of reinfection (Fig. 2C). Confirmation of reinfection by either reverse transcription polymerase chain reaction (RT-PCR) ($n = 8$) or symptomatic illness ($n = 2$) was available for 10 of the 25 subjects. No severe illness or hospitalization was reported in reinfections.

Time periods of increased transmission were associated with declining RT-PCR cycle threshold (Ct) values (Fig. 3A and fig. S1), attributable to a higher proportion of recently infected individuals with high viral loads (*6*). However, the Ct decline was far greater in April 2021 (dCt, SE: −4.06, 0.27; $P < 0.001$) than seen previously (November 2020 dCt: ~−1.5). Comparing April 2021 with November 2020, high viral load samples (Ct < 20) doubled in clinical samples ($P < 0.001$) and nearly doubled in campus surveillance testing data, where most positives were from recently infected individuals [15% ($n = 297$) versus 9% ($n = 358$); $P = 0.02$].

Genome analysis trends, in representative samples drawn from the general population over the same period, showed seeding and expansion of B.1.1.7 (Alpha), B.1.617.1 (Kappa), and B.1.617.2 (Delta) lineages in 2021, with

---

[1] National Centre for Disease Control, Delhi, India. [2] CSIR-Institute of Genomics and Integrative Biology, New Delhi, India. [3] Academy for Scientific and Innovative Research, Ghaziabad, India. [4] Medical Research Council (MRC) Centre for Global Infectious Disease Analysis, Jameel Institute, School of Public Health, Imperial College London, London, UK. [5] Department of Mathematics, Imperial College London, London, UK. [6] Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. [7] Department of Medicine, Cambridge Institute of Therapeutic Immunology and Infectious Disease (CITIID), University of Cambridge, Cambridge, UK. [8] Africa Health Research Institute, KwaZulu-Natal, South Africa.

***Corresponding author. Email: partho_rakshit@yahoo.com (P.R.); a.agrawal@igib.in (A.A.)**

†These authors contributed equally to this work.

‡The Indian SARS-CoV-2 Genomics Consortium (INSACOG) collaborators and affiliations are listed in the supplementary materials.

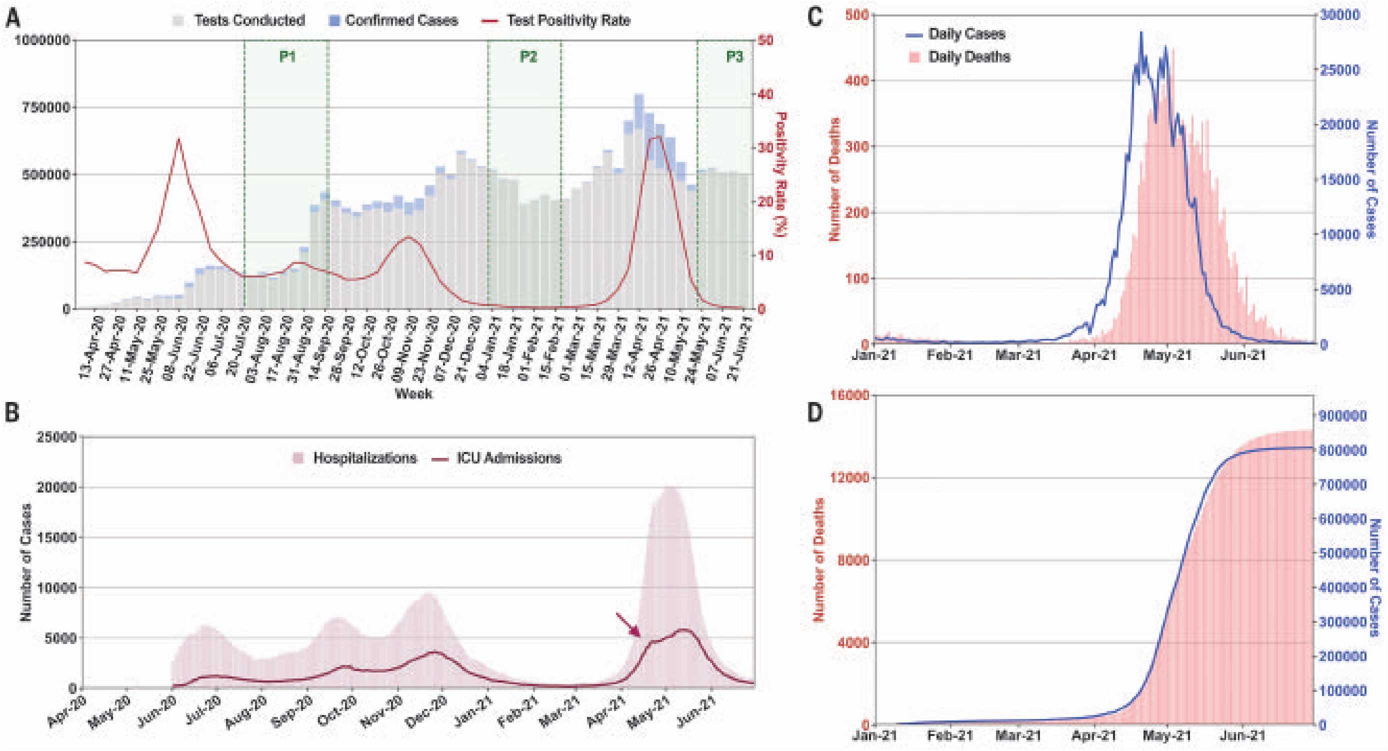


**Fig. 1. Multiple surges of SARS-CoV-2 infections in Delhi with an overwhelming outbreak in April and May 2021.** (**A**) Weekly tests, confirmed cases, and test positivity rate in Delhi from April 2020 to June 2021. Sample collection period for CSIR serosurveys is marked as P1 to P3. (**B**) Number of hospitalizations and ICU admissions plotted on a daily basis from June 2020 to June 2021. The arrow marks the possible saturation of ICU capacity (*3*). (**C**) Daily cases and daily deaths from January 2021 to June 2021. (**D**) Time-advanced and scaled cumulative cases, fitted to cumulative deaths. Time advancement of cumulative reported cases by 8 days was done for maximal coincidence with scaled cumulative deaths. Case fatality ratio = averaged scaling factor (cumulative deaths / time-advanced cumulative cases). Mean ± SD, 0.019 ± 0.003.

Delta becoming the dominant lineage in Delhi during April (Fig. 3B). The proportion of the Delta variant was strongly correlated to the rise in cases and health care stress (fig. S2). Overall, the genomic and epidemiological data were most consistent with the hypothesis that a new variant with higher infectivity, Delta, was driving the unexpected overwhelming surge in Delhi. Recent in vitro data supports the possibility of a higher replication rate for Delta, thereby explaining potentially higher viral loads in RT-PCR data and greater transmissibility (*7*).

We further investigated the sequence of seeding and spread of the Alpha, Kappa, and Delta variants of SARS-CoV-2. Phylogenetic analysis showed common origins between the Alpha variants in Delhi and Punjab, and between the Kappa or Delta variants in Delhi and Maharashtra, where Kappa and Delta were first sequenced (Fig. 3C and fig. S3). There was a recurring pattern of initial smaller outbreaks with Alpha, followed by larger outbreaks coinciding with Alpha-to-Delta transition across all neighboring states (fig. S4). The substantial relative growth advantage of Delta was explored in terms of transmissibility and/or immune escape. The rise of Delta, but not other lineages, was temporally coincident with a rise in the test positivity rate (TPR) and new cases during the surge (fig. S5). While overall vaccination levels were only about 5% in Delhi, most health care workers had received one or two doses of ChAdOx1-nCov19 (AstraZeneca/Serum Institute, India) or BBV152 (Bharat Biotech, India) (*8*, *9*). We sequenced 24 breakthrough infections starting at least 1 week after the first dose, collected between 22 March and 28 April 2021 at the National Centre for Disease Control (NCDC). The ratio of Delta to non-Delta lineages was 850:1211 from 20 March to 30 April. In contrast, the ratio of Delta to non-Delta lineages was 13:3 in 16 breakthroughs after one dose and 7:1 in eight breakthroughs after the second dose of vaccine. Although the small sample size and lack of a formal control group preclude definitive analysis, estimated higher odds for Delta in vaccination breakthroughs (odds ratio, 7.1; 95% CI, 2.4 to 20.9) corroborate other reports of reduced vaccine effectiveness against Delta (*10*).

To better characterize how the properties of Delta might differ from other SARS-CoV-2 lineages previously circulating in the city, we used a Bayesian model of SARS-CoV-2 transmission and mortality that simultaneously models the dynamics of two categories of virus (B.1.617.2 and non-B.1.617.2) (*11*), while also explicitly incorporating natural waning of immunity derived from prior infection, with the duration of immunity consistent with the results of recent longitudinal cohort studies (*12*, *13*). Details of the model and input data are given in the supplementary methods. Briefly, the model is fitted to COVID-19 mortality data, genomic sequence data presented here and from GISAID [with Phylogenetic Assignment of Named Global Outbreak Lineages ( PANGOLIN) classification] (*14*–*16*), and serological data presented alongside an additional longitudinal serosurvey carried out in the city from July to December 2020 (*17*). Substantial uncertainties remain as to the date of introduction of B.1.617.2 into Delhi and the degree of COVID-19 death underascertainment. We therefore explored a range of different scenarios in which we varied underascertainment (10, 33, 50, and 66%) and introduction dates (15 January 2021, 31 January 2021, 14 February 2021, and 28 February 2021).

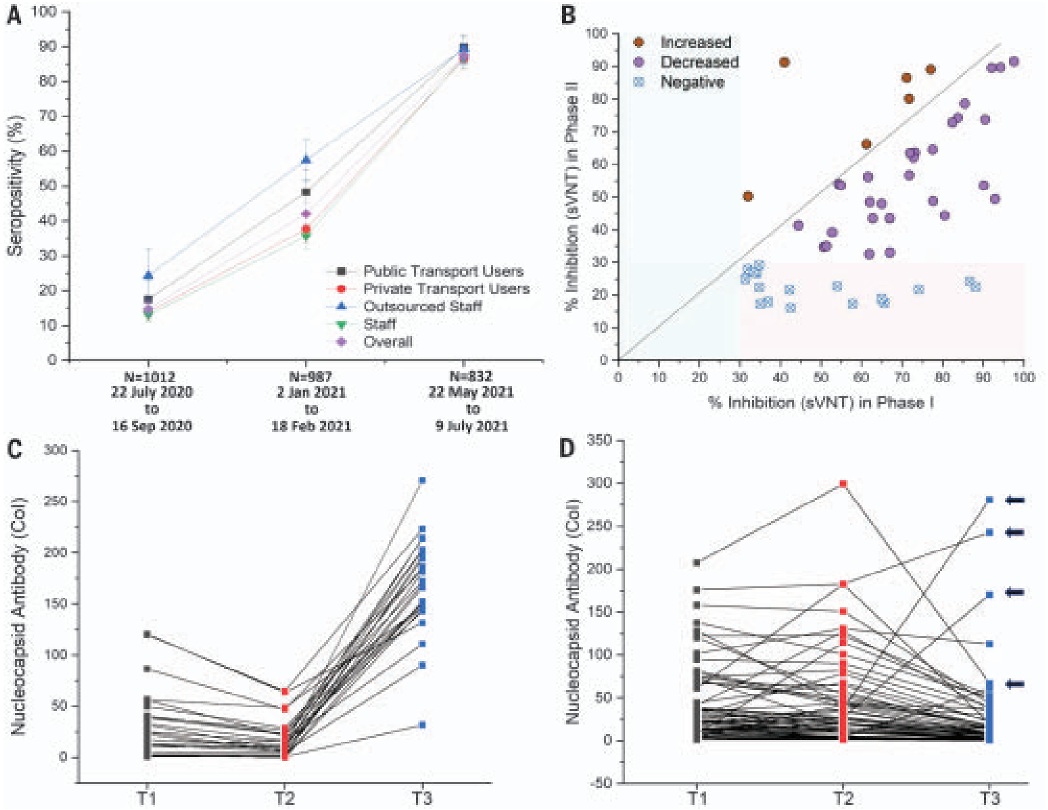


**Fig. 2. Serological estimates of prior infections, preexisting immunity, and new infections for the April and May 2021 outbreak.** (**A**) Seropositivity in the CSIR cohort, subdivided by nature of employment and use of public transport, plotted for different time periods (phase I to phase III, proportion ± 95% CI). Details are provided in table S1. (**B**) Variability and temporal decline in neutralization capacity estimated by sVNT assay between phases I and II ($n = 52$ subjects). (**C**) Serial antibody concentration measurements in initially seropositive subjects ($n = 91$). Pattern suggestive of reinfections is shown (decline followed by rise, $n = 25$). (**D**) Remaining data ($n = 66$ subjects), with four indeterminate reinfection cases indicated with arrows. Antibody concentration is reported in multiples of the assay cutoff index value (CoI).

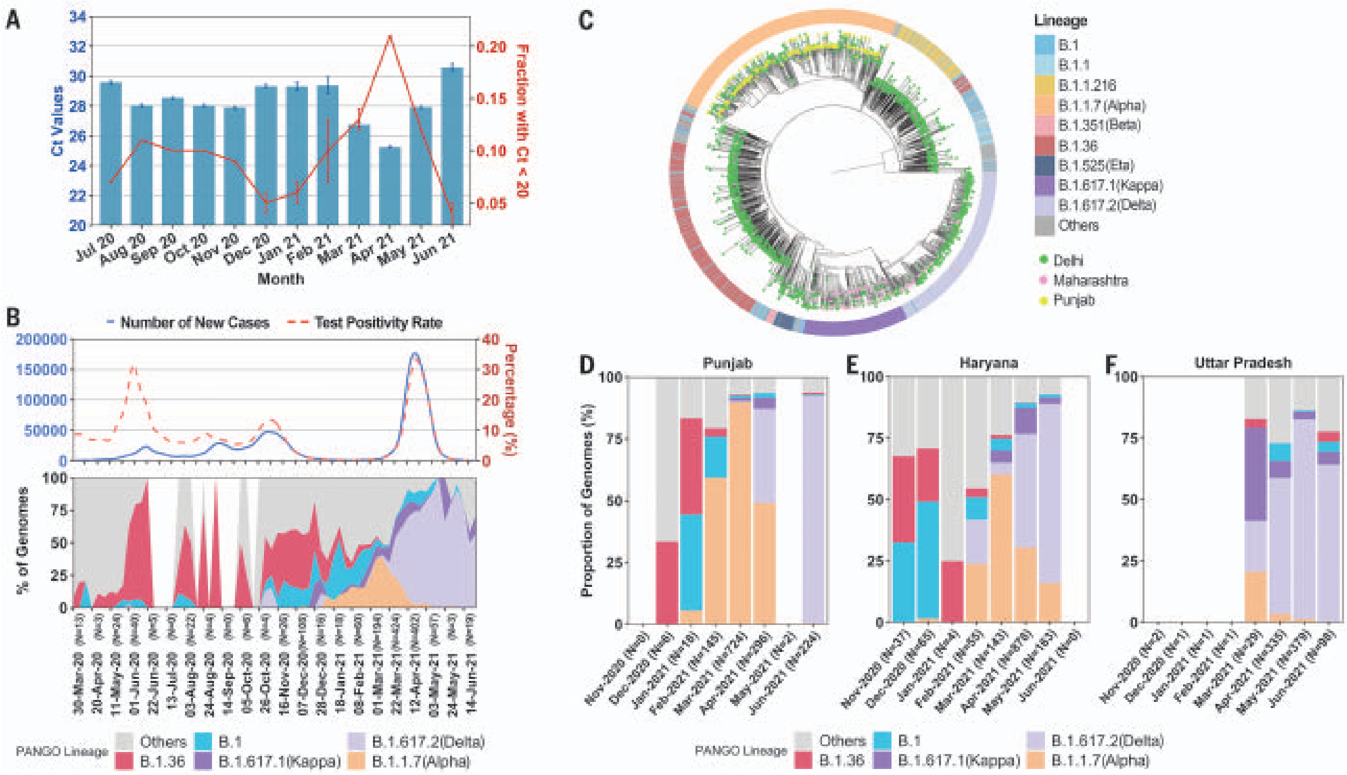


**Fig. 3. Genomic-epidemiologic correlations.** (**A**) Time trends of Ct values (mean ± SE) and high viral load samples (proportion ± SE) for *Orf1* gene (E gene data, fig. S1). (**B**) Smoothed graph of main lineages in Delhi from March 2020 to May 2021 in biweekly increments. New cases and TPR are aligned and plotted on the same timeline. (**C**) Phylogenetic analysis for variant of concern (VOC) strains between Delhi and states (Punjab and Maharashtra) with known VOC outbreaks before April 2021. Further analysis suggesting a superspreading event for Alpha is shown in fig. S3. (**D** to **F**) Month-wise proportions of different lineages ($n > 3$ sequenced samples) in states surrounding Delhi. Additional data are shown in figs. S4 and S5.

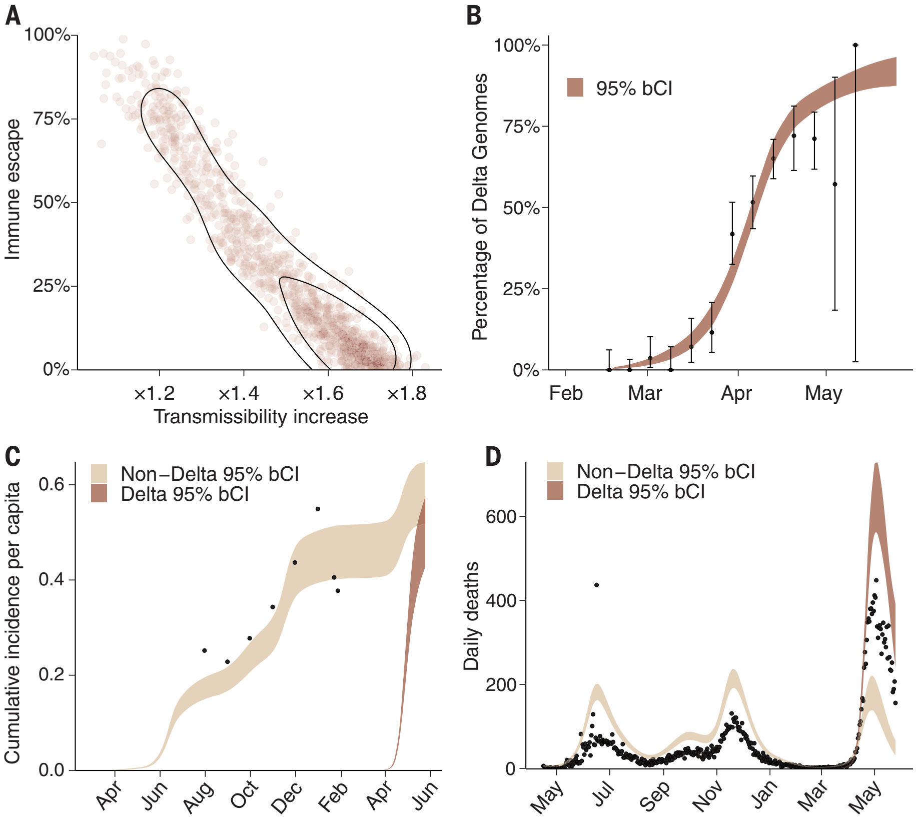


**Fig. 4. Estimates of the epidemiological characteristics of the Delta variant.** Values were inferred from a two-category Bayesian transmission model fitted to mortality, serosurvey, and genomic data from Delhi, India. (**A**) Joint posterior distribution, with isoclines corresponding to the 90% and 50% enclosures of posterior density of the Delta variant immune escape and transmissibility increase relative to non-Delta categories. Immune escape has a median of 20% with 50% Bayesian credible interval (bCI) of 10 to 50%, and transmissibility has a median increase of 1.5 with 50% bCI of 1.3 to 1.7. (**B**) Delta fraction over time, inferred by the model. Black dots represent genome sampling data points, with exact binomial confidence intervals. (**C**) Serosurvey data (black dots) and inferred cumulative incidence for Delta and non-Delta variant categories. (**D**) Mortality data (black dots) and inferred deaths assuming 50% underreporting. Other underascertainment scenarios are presented in the supplementary materials.

Using this framework for an introduction date of 14 February 2021 and mortality underascertainment of 50%, our results, shown in Fig. 4, indicate that the Delta variant is 1.3- to 1.7-fold [50% Bayesian credible interval (bCI)] more transmissible than earlier and co-circulating SARS-CoV-2 lineages in Delhi, including the Alpha variant. Notably, the model also indicates that the Delta variant can partially evade immunity elicited by prior infection, with prior infection providing only 50 to 90% (50% bCI) of the protection against infection with the Delta variant that it provides against previous lineages. There is an inherent trade-off between transmissibility and immune escape, and the worst-case scenario of both very high transmissibility and immune escape is rejected a posteriori by the data. Figure 4A also highlights the nature of uncertainty in the exact level of immune escape and transmissibility increase, because these inferred characteristics of the Delta variant are collinear given the modeling framework used and data currently available. The main limitations of the model are due to biases in the data and our choice of priors. For example, there is an unknown degree of underreporting, and serological estimates are likely to be systematically biased. Similarly, there are uncertainties in estimates of the temporal waning of immunity, the date of the first introduction of Delta, and the true infection fatality rate of SARS-CoV-2 variants in Delhi. We explicitly consider these limitations and partially mitigate such biases, as described in the supplementary materials. Overall, the main findings are robust to variation in prior assumptions, including both changing underascertainment and the date of introduction (supplementary methods and tables S2 and S3). The results are valid for a population where most of the immunity arose from prior infection (rather than vaccination), which is true for Delhi. On the basis of median estimates of the model (Fig. 4A) and high transmissibility of the background Alpha lineage (*18*), Delta should potentially be at least twice as transmissible as the wild-type lineage.

We note that after the massive Delta outbreak, new cases in Delhi and other North Indian states have stayed extremely low, with TPR in Delhi below 0.1% as of September 2021. This fits the serological picture presented in Fig. 2, with a very high fraction of the population being recently infected and with good immunity to Delta. However, Delta outbreaks have continued in parts of India and elsewhere in the world, despite moderately high seropositivity or vaccination levels that were previously considered to be adequate (*7*, *19*, *20*). We conclude that the Delta variant is capable of initiating fast-rising outbreaks in populations with immune responses to prior variants, resulting in reinfections and vaccination breakthroughs. Public health strategies may need to be revised to account for variants with heightened transmissibility and immune escape.

## REFERENCES AND NOTES

1. M. V. Murhekar *et al.*, *Int. J. Infect. Dis.* **108**, 145–155 (2021).
2. S. Jain (@SatyendarJain), Twitter post, 2 February 2021, 5:19 a.m.; https://twitter.com/satyendarjain/status/1356547817427226624?lang=en.
3. Government of National Capital Territory of Delhi, Orders F.23/Misc/COVID-19/DGHS/NHC/2020/Pt XVIII/90-97 and No.01/H&FW/COVID-Cell/2020/07/ss4hfw/726, 12 April 2021; http://health.delhigovt.nic.in/wps/wcm/connect/doit_health/Health/Home/Covid19/Covid+19+Related+order+April+2021.
4. S. Naushin *et al.*, *eLife* **10**, e66537 (2021).
5. J. M. Dan *et al.*, *Science* **371**, eabf4063 (2021).
6. J. A. Hay *et al.*, *Science* **373**, eabh0635 (2021).
7. P. Mlcochova *et al.*, *Nature* 10.1038/s41586-021-03944-y (2021).
8. P. M. Folegatti *et al.*, *Lancet* **396**, 467–478 (2020).
9. R. Ella *et al.*, medRxiv 2021.06.30.21259439 [Preprint] (2021). https://doi.org/10.1101/2021.06.30.21259439.
10. J. Lopez Bernal *et al.*, *N. Engl. J. Med.* **385**, 585–594 (2021).
11. S. Flaxman *et al.*, *Nature* **584**, 257–261 (2020).
12. C. H. Hansen, D. Michlmayr, S. M. Gubbels, K. Mølbak, S. Ethelberg, *Lancet* **397**, 1204–1212 (2021).
13. V. J. Hall *et al.*, *Lancet* **397**, 1459–1469 (2021).
14. Y. Shu, J. McCauley, *Euro Surveill.* **22**, 30494 (2017).
15. A. Rambaut *et al.*, *Nat. Microbiol.* **5**, 1403–1407 (2020).
16. Á. O'Toole *et al.*, *Virus Evol.* **7**, veab064 (2021).
17. A. Velumani *et al.*, medRxiv 2021.03.19.21253429 [Preprint] (2021). https://doi.org/10.1101/2021.03.19.21253429.
18. N. G. Davies *et al.*, *Science* **372**, eabg3055 (2021).
19. D. Planas *et al.*, *Nature* **596**, 276–280 (2021).
20. S. Mishra *et al.*, *EClinicalMedicine* **39**, 101064 (2021).
21. banijolly/ncov-Delhi-Epidemiology: nCoV Delhi-Epidemiology, version 1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.5521073.
22. T. Mellan, M. Swapnil, C. Whittaker, S. Bhatt, S. Flaxman, ImperialCollegeLondon/Delta_Variant_Delhi: Genomic characterization and Epidemiology of an emerging SARS- CoV-2

variant in Delhi, India, version 1.0, Zenodo (2021); https://doi.org/10.5281/zenodo.5524442.

**ACKNOWLEDGMENTS**

Suggestions by N. K Arora, G. Padmanabhan, and members of the Scientific and Clinical Advisory Group of INSACOG are acknowledged. Coordinators of CSIR labs and centers are acknowledged for CSIR-Cohort. We gratefully acknowledge the authors from the originating laboratories and the submitting laboratories, who generated and shared via GISAID genetic sequence data on which this research is based. A full list acknowledging GISAID contributors for the data used in this study can be found in data S3. **Funding:** Support from the Ministry of Health and Family Welfare, the Council of Scientific and Industrial Research (CSIR-Phenome India Grant, MLP-2007; InGen-CoV2, MLP-2005), and the Department of Biotechnology (INSACOG), India, is acknowledged. **Author contributions:** Conceived of the study: A.A., P.R., S.K.S., R.K.G., S.Ka., S.S., M.F., D.D., R.P., S.B., S.Fl., R.C.B., and B.J. Designed the study and experiments: A.A., P.R., S.K.S., R.K.G., S.Ka., S.S., M.F., R.P., U.S., P.Si., M.D., R.C.B., K.P., R.V.S., R.Mar., and M.S.D. Performed experiments: M.S.D., R.Mar., R.V.S., R.C.B., S.N., M.D., P.Si., U.S., R.U., M.K.D., M.K.S., M.I., V.Se., and P.Sh. Collected and analyzed patient data: M.S.D., R.Mar., R.V.S., K.P., M.D., P.Si., U.S., M.K.S., U.C., N.S., H.L., H.G., P.Ma., S.Ku., H.C., and P.R. Performed bioinformatics analyses: M.S.D., R.Mar., R.V.S., K.P., B.J., S.Fa., N.B., R.Mau., N.J., P.Me., V.A., A.V.R., and L.T. Performed statistical analyses: A.A., V.Sa., T.A.M., S.M., C.W., S.Fl., S.B., and D.D. Interpreted data: A.A., P.R., M.S.D., R.Mar., R.V.S., K.P., B.J., V.Sa., S.N., M.R., T.A.M., S.M., C.W., R.U., N.B., L.T., S.Fl., S.B., R.P., D.D., M.F., S.S., S.Ka., R.K.G., and S.K.S. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data used in the analysis are accessible in the supplementary materials as data files or in Zenodo (*21*), including accession numbers for all genomic data that have been deposited at GISAID. Code and data for replication of the Bayesian inference are available in Zenodo (*22*). Biological samples from infectious patients and patient details were collected for the purpose of public health and cannot be shared for reasons of confidentiality and to comply with Indian government regulations. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

**SUPPLEMENTARY MATERIALS**

science.org/doi/10.1126/science.abj9932
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S3
List of INSACOG Consortium Members
References (*23*–*42*)
MDAR Reproducibility Checklist
Data S1 to S9

15 June 2021; accepted 6 October 2021
Published online 14 October 2021
10.1126/science.abj9932

---

**DENGUE EVOLUTION**

# Antigenic evolution of dengue viruses over 20 years

**Leah C. Katzelnick[1,2], Ana Coello Escoto[1,2], Angkana T. Huang[1,3], Bernardo Garcia-Carreras[1], Nayeem Chowdhury[1], Irina Maljkovic Berry[4], Chris Chavez[1], Philippe Buchy[5], Veasna Duong[6], Philippe Dussart[6], Gregory Gromowski[4], Louis Macareo[3], Butsaya Thaisomboonsuk[3], Stefan Fernandez[3], Derek J. Smith[7], Richard Jarman[4], Stephen S. Whitehead[8]*, Henrik Salje[1,9]*, Derek A. T. Cummings[1]***

**Infection with one of dengue viruses 1 to 4 (DENV1-4) induces protective antibodies against homotypic infection. However, a notable feature of dengue viruses is the ability to use preexisting heterotypic antibodies to infect Fcγ receptor–bearing immune cells, leading to higher viral load and immunopathological events that augment disease. We tracked the antigenic dynamics of each DENV serotype by using 1944 sequenced isolates from Bangkok, Thailand, between 1994 and 2014 (348 strains), in comparison with regional and global DENV antigenic diversity (64 strains). Over the course of 20 years, the Thailand DENV serotypes gradually evolved away from one another. However, for brief periods, the serotypes increased in similarity, with corresponding changes in epidemic magnitude. Antigenic evolution within a genotype involved a trade-off between two types of antigenic change (within-serotype and between-serotype), whereas genotype replacement resulted in antigenic change away from all serotypes. These findings provide insights into theorized dynamics in antigenic evolution.**

Antigenic evolution occurs in many viruses. Viral proteins recognized by the immune system change, enabling evasion of host immunity induced by prior infection with similar viruses (*1*). Dengue viruses 1 to 4 (DENV1-4) provide an example of how preexisting heterotypic antibodies may not only be evaded but also exploited by the virus to aid infection by facilitating entry into and replication in immune cells, thus inducing higher viral load and immunopathological events that increase disease severity (*2*–*6*). DENV1-4 are responsible for ~100 million infections, 50 million febrile dengue cases, ~500,000 hospitalizations, and 10,000 to 25,000 deaths annually (*7*). High levels of cross-serotype reactive antibodies can protect against secondary DENV infection with a different serotype, whereas low to intermediate levels increase risk of dengue hemorrhagic fever or dengue shock syndrome as a result of antibody-dependent enhancement (*3*–*6*).

DENV1-4 vary antigenically within and between serotypes, but strong evidence for antigenic escape has not been found. Each DENV serotype consists of four to seven genotypes that differ from one another by ≤10% at the amino acid level across the envelope protein (*8*). Antibodies from naturally infected and vaccinated individuals differentially neutralize distinct genotypes and even distinct clinical isolates (strains) of each serotype (*9*–*12*). Homotypic immunity is generally protective, although protection against clinical disease is not always complete against viruses of a different genotype; this could also potentially reduce vaccine efficacy (*13*, *14*). Further, there are large differences in antigenic similarity between a given DENV strain and strains of different DENV serotypes (*9*, *15*). Phylogenetic analyses indicate that genotype replacement events, defined as when a previously common viral lineage vanishes in a given location and a related but distinct lineage becomes dominant, may be driven by natural selection and immune pressure, although population bottlenecks are an alternative explanation (*16*–*18*). However, few studies have linked genetic or antigenic differences between strains to epidemic magnitude and severity (*17*, *19*, *20*). Mechanistic transmission models and empirical observations suggest that temporary cross-protection and antibody-dependent enhancement—including enhanced probability of disease, infectivity, or susceptibility to secondary infections—along with other spatial, temporal, and vector-associated parameters are drivers of DENV epidemic dynamics (*18*, *21*–*27*). However, it remains controversial whether antigenic variation observed among DENV1-4 is biologically relevant and associated with epidemic dynamics of cocirculating strains. If DENV1-4 are evolving antigenically, changes are expected to be most evident in a single highly endemic geographic

---

[1]Department of Biology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA. [2]Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [3]Department of Virology, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand. [4]Viral Diseases Branch, Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA. [5]GlaxoSmithKline (GSK) Vaccines, 637421 Singapore, Singapore. [6]Institut Pasteur in Cambodia, Réseau International des Instituts Pasteur, Phnom Penh 12201, Cambodia. [7]Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ UK. [8]Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [9]Department of Genetics, University of Cambridge, Cambridge, CB2 3EJ UK.
***Corresponding authors. Email: swhitehead@niaid.nih.gov (S.S.W.); hs743@cam.ac.uk (H.S.); datc@ufl.edu (D.A.T.C.)**

location where strains interact directly with immunity derived from other currently or previously circulating strains.

In this study, we tested whether DENV1-4 circulating in Bangkok, Thailand, changed antigenically over two decades in relation to each other and a selection of globally representative DENV1-4 strains. We conducted full-genome sequencing of 1944 clinical DENV isolates systematically sampled between 1994 and 2014 at the Queen Sirikit National Institute of Child Health (QSNICH). QSNICH is a tertiary children's hospital in central Bangkok, Thailand, that serves as the city's main referral center for cases of dengue requiring hospitalization (*28*, *29*). DENV1-4 have circulated in Bangkok since at least 1962; transmission is observed annually and year-round, with larger epidemics occurring periodically (*29*–*31*). Between 1994 and 2014, each serotype was dominated by a single genotype: genotype I for DENV1, genotype Asian I for DENV2, genotype II for DENV3, and genotype I for DENV4 (*32*, *33*) (fig. S1). A different DENV4 genotype was found circulating at lower levels, and single representatives of distinct DENV2 genotypes were isolated. DENV3 was the only serotype to undergo a genotype replacement event during this period, with genotype III displacing genotype II.

We systematically selected 348 of the sequenced Thailand DENV1-4 strains for antigenic characterization. Sampling was balanced across years and included representation of amino acid variation in the envelope (E) and premembrane (prM) proteins to increase the likelihood of capturing any antigenic change (Fig. 1, A to E, fig. S1, and table S1). To place the Thailand DENV1-4 in a temporal, regional, and global context, we also antigenically characterized strains from 20 different countries isolated between 1944 and 2012 ($n$ = 64). All strains ($n$ = 412 total) were characterized by a plaque (immunofocus) reduction neutralization test with a 50% neutralization cutoff ($\mathrm{PRNT}_{50}$), the gold standard method for measuring serological immunity to DENV1-4. High PRNT titers are correlated with vaccine efficacy and reduced risk of dengue; lower PRNT titers may be associated with increased risk of severe dengue (*4*, *34*). All viruses were titrated by PRNT on mosquito cells (*Aedes albopictus* C6/36) against a panel of antisera from 20 African green monkeys (*Chlorocebus sabaeus*) ($n$ = 7957 titrations) (Fig. 1F, table S2, and methods), each inoculated with a distinct DENV strain. This panel was used previously to characterize DENV antigenic diversity (*9*, *35*). Sera collected 90 days after inoculation had high $\mathrm{PRNT}_{50}$ titers and closely approximated the long-term antibody response (day 150) (*9*). $\mathrm{PRNT}_{50}$ titers were adjusted by virus to control for experimental conditions that systematically modified $\mathrm{PRNT}_{50}$ titers, regardless of sampling year or serotype: (i) duration of virus serum incubation and (ii) realiquoting of old virus stocks versus reamplification immediately before titration. We estimated the antigenic relationships among strains by antigenic cartography, a method that converts serum $\mathrm{PRNT}_{50}$ titers into units of antigenic distance by expressing titer data in maps of reduced dimensions compared with the full data (*1*, *9*). We used cross-validation (100 maps, each with a random 75% of titers) to identify the coordinates and

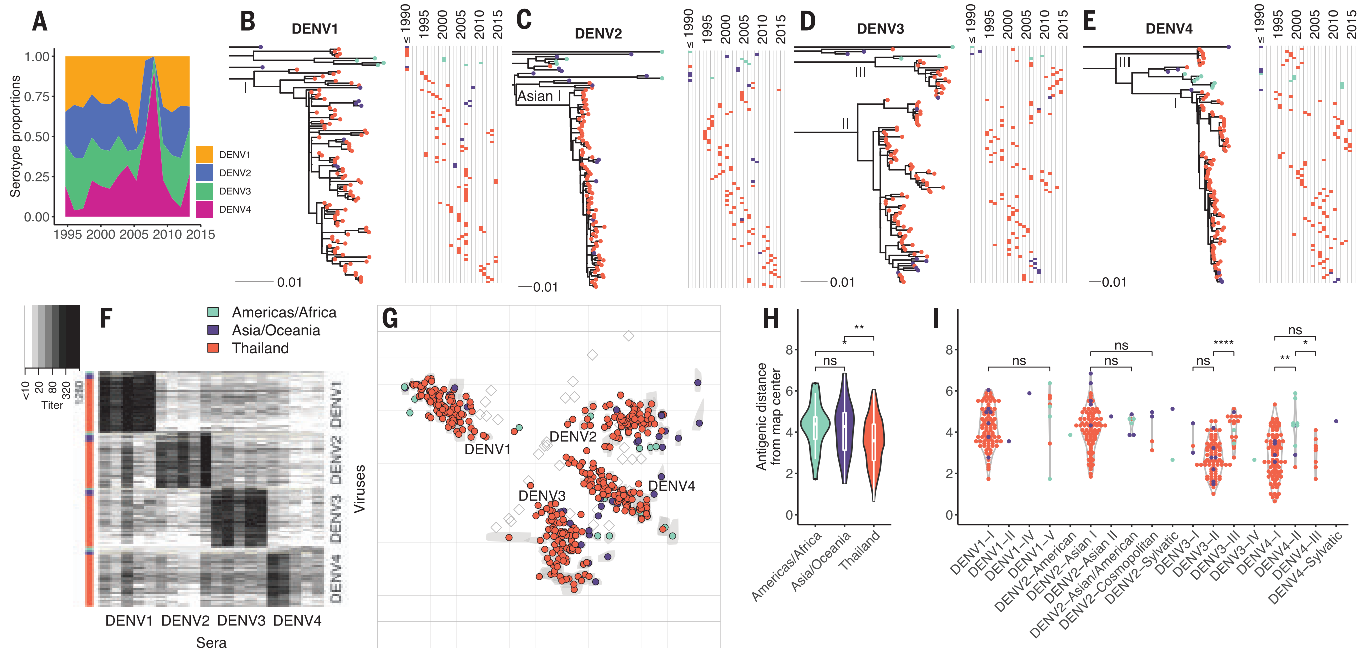


**Fig. 1. Genetic and antigenic characteristics of DENV1-4 strains isolated in Bangkok, Thailand, in relation to global DENV strains.** (**A**) Proportion by serotype of 1944 clinical DENV strains isolated between 1994 and 2014 at QSNICH. Strains for antigenic characterization were selected from this full set. (**B** to **E**) Evolutionary relatedness among E protein sequences of DENV1-4 ($n$ = 348) from Thailand (1994 to 2014) compared with strains from other countries or periods in time ($n$ = 64). Maximum likelihood phylogenetic trees were built by generalized time reversible nucleotide substitution models with gamma-distributed rate variation and allowing for invariant sites (GTR+G4+I). Strains are colored to indicate the geographic area where each was isolated (Americas and Africa, Asia and Oceania, or Thailand). Corresponding time series show the years of strain isolation. (**F**) Heat map of $\mathrm{PRNT}_{50}$ titers ($n$ = 7957) for all DENV1 ($n$ = 105), DENV2 ($n$ = 99), DENV3 ($n$ = 103) and DENV4 ($n$ = 105) strains titrated against antisera from nonhuman primates ($n$ = 20, 5 per serotype) each inoculated with a genetically distinct global DENV strain. Rows correspond to DENV strains (row colors indicate region) and columns correspond to antisera. (**G**) Antigenic map made in two dimensions of all DENV1-4 strains. Gray shapes indicate interquartile range of coordinates for each virus based on cross-validation maps. Colored circles correspond to median coordinates for each virus. Each grid square side corresponds to a twofold dilution in the $\mathrm{PRNT}_{50}$ assay, and distance is interpretable in any direction. Sera are represented as open squares. Violin plots of antigenic distances of each virus from the center of the 3D antigenic map by location of virus isolation (**H**) or genotype (**I**). Global significance tests were conducted by a Kruskal-Wallis rank-sum test, followed by pairwise comparisons by the Wilcoxon rank-sum test. ns, not significant, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

map dimensions (exploring 2 to 10 dimensions) for which map antigenic distances most closely predicted excluded titers (fig. S2). All antigenic analyses were performed on 3D maps, which were found to optimally represent the titer data (data S1). 2D maps had similar performance and are shown as well.

On the 2D and 3D antigenic maps, extensive antigenic diversity was observed within serotypes (Fig. 1G, 95th percentile of within-serotype distances in a 3D map, 8.9 to 19.2-fold difference in PRNT titer, data S1). As we have observed previously, some strains were as close antigenically to strains of other serotypes as to strains of the same serotype (*9*). DENV1-4 strains from Thailand were closer antigenically to each other (measured as distance from the map center) than strains from other countries in Asia and Oceania or the Americas and Africa (Fig. 1H). This effect remained after accounting for the greater number of Thailand strains in the dataset (fig. S3). These differences were explained in part by the genotypes circulating in Thailand. The dominant genotypes of DENV4 circulating in Thailand, genotypes I and III, were significantly more antigenically central on the maps than DENV4 genotype II strains, which circulate in the Americas and Africa (Fig. 1I). DENV3 genotype II strains were also significantly more similar antigenically to other serotypes than DENV3 genotype III strains (Fig. 1I), which became the dominant genotype in Thailand at the end of the observational period (Fig. 1D).

Thailand DENV1-4 strains appeared to be evolving antigenically over time (Fig. 2 and movie S1). To measure these antigenic dynamics quantitatively, we fitted antigenic distances from the 3D map as a function of time by using linear regression and generalized additive models (GAMs) with 100 bootstrap resamples to construct confidence intervals (CIs) (table S3). Between 1994 and 2014, the serotypes moved away from the center of the antigenic map, dropping in neutralization by 40% overall (95% CI: 20 to 53%) (Fig. 3A). DENV1, DENV3, and DENV4 each moved away from other serotypes, with cross-neutralization dropping by 52% for DENV1, 74% for DENV3, and 39% for DENV4 over 20 years (Fig. 3A, vertical plots). Both the raw data (means and standard deviations of annual antigenic data) and GAM fits reveal nonlinear antigenic dynamics [estimated degrees of freedom (EDF)]: 4.1, $P$ <0.001, Fig. 3A), with antigenic distance fluctuating around an overall increase in time; this effect was observed individually for DENV1, DENV3, and DENV4 (Fig. 3A, vertical plots; detrended antigenic dynamics are shown in fig. S4). The serotypes also moved away from one another, with an average decrease in neutralization between pairwise serotypes of 65% over the 20-year period (Fig. 3B). Again, the distance between serotypes fluctuated (EDF: 8.5, $P$ <0.001) from as little as 4.9 antigenic units (29-fold difference in PRNT$_{50}$ titers) up to 6.9 antigenic units (120-fold difference) (Fig. 3B).

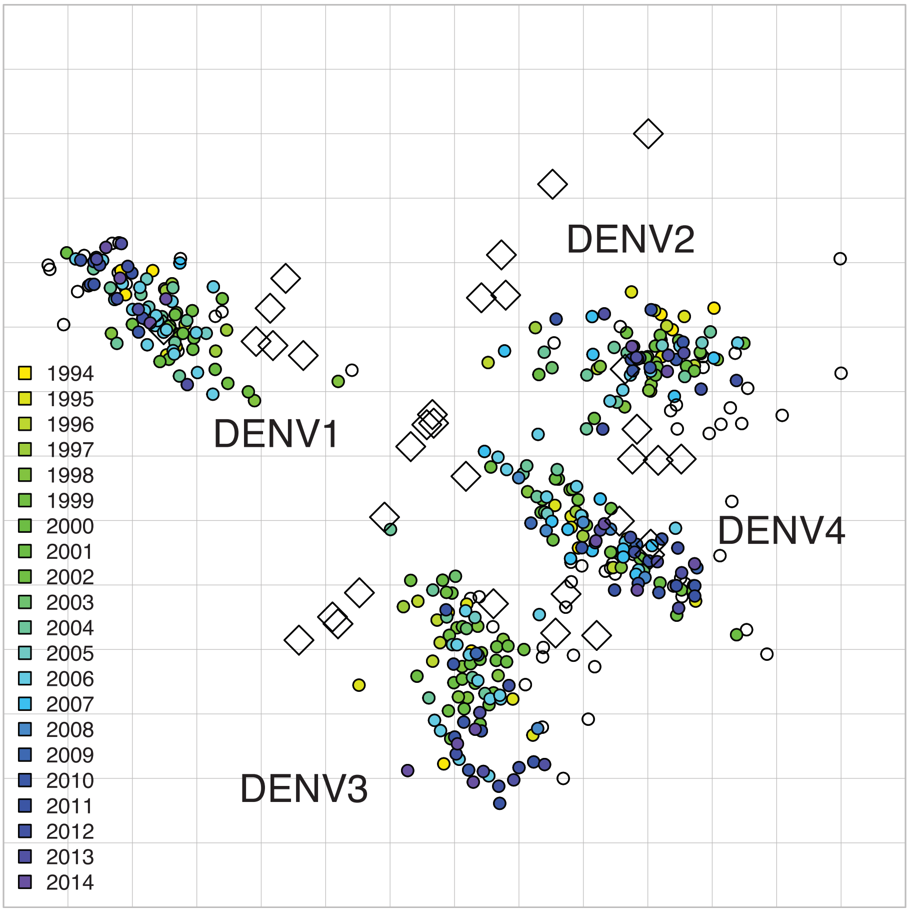


**Fig. 2. 2D antigenic map of Thailand DENV1-4 colored by year of isolation.** Open circles show global viruses; open squares show the serum positions. Serotype clusters are labeled. Each grid square side in both dimensions is equivalent to a twofold dilution in the PRNT$_{50}$ assay.

The observed increase in antigenic distance between DENV serotypes is consistent with the hypothesis that mounting immunity to previously circulating strains selects for viruses that are antigenically different (*1*). Homotypic DENV immunity is potently neutralizing and long lasting, and thus major epidemics may select for antigenic differences relative to previously circulating strains of the same serotype. Further, high cross-serotype immunity induced in the first months after primary DENV infection or for years after secondary DENV infection may select for antigenic evasion of heterotypic immunity after large epidemics of other serotypes. However, we also observed that periodically, the serotypes evolved to be more antigenically similar. One hypothesis is that cross-serotype antibodies wane after primary infection to titers that mediate antibody-dependent enhancement of infection, viral load, and severity, "pulling" a serotype antigenically toward other serotypes (*2*). Alternatively, structural constraints on protein function imposed by the need for DENV to efficiently replicate both in the human host and mosquito vector may limit the mode of antigenic change possible for DENV at a given time. For instance, to evade homotypic immunity, strains may change to resemble other serotypes when heterologous neutralization is a weak selective pressure. If this is true, we would expect to observe an inverse relationship between homotypic and heterotypic antigenic dynamics and a link between homotypic immune evasion and larger epidemics. It is also possible that within and between serotypes, antigenic evolution proceeds independently if the epitopes targeted are distinct (*36*).

We tested the hypothesis that within and between serotypes, antigenic change is correlated. Within-serotype antigenic change was measured as the pairwise antigenic distance between each strain and the 1994 and 1995 strains of the matched serotype (Fig. 3C). On average, the serotypes became more antigenically distinct from earlier strains of the

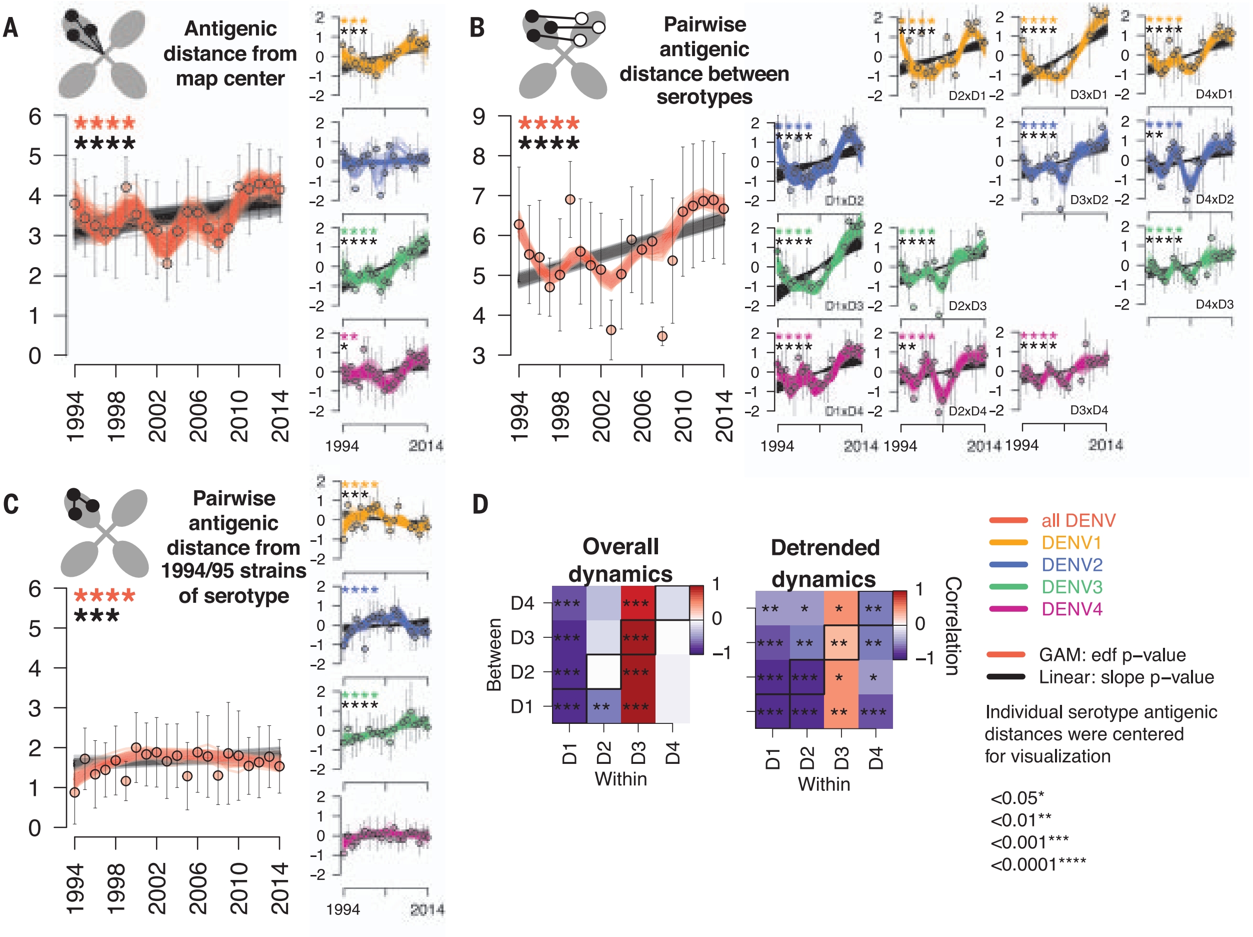


**Fig. 3. Antigenic dynamics of Thailand DENV1-4 strains isolated between 1994 and 2014.** Antigenic change over time for all serotypes, measured as (**A**) distance from the map center, (**B**) pairwise distance between serotypes each year, and (**C**) pairwise distance from 1994 to 1995 strains of the same serotype. Illustrations of antigenic maps in the top left corner of each plot depict each antigenic distance metric. For each antigenic time series, antigenic distances were bootstrap sampled and used to construct 100 linear (black lines) and 100 nonlinear generalized additive (colored lines) models. Mean and standard deviation of antigenic distances are shown as colored circles with black error bars, respectively. Black stars indicate significant linear change (slope) and colored stars indicate statistically significant nonlinearity (effective degrees of freedom). Models were run for all serotypes combined (with a variable to adjust for serotype, large plots, and the *y* axis showing measured distances) and for each serotype separately (vertical plots and *y* axis showing distances centered at 0 to facilitate comparison of relative change across plots). (**D**) Pearson's correlation coefficients and corresponding p-values of bootstrapped (*n* = 1000) within serotype (columns) versus between serotype (rows) overall antigenic dynamics (no adjustment) and detrended dynamics (linear model subtracted from the GAM before analysis). Diagonal black boxes correspond to distance from the map center and off diagonal indicates pairwise distance between serotypes. Color indicates correlation (range −1 to 1) and significance is indicated by stars.

same serotype before gradually switching back, with oscillations in antigenic distance throughout the period (Fig. 3C). These dynamics were significant for DENV1 and DENV2 (Fig. 3C, vertical plots). We then measured the Pearson's correlation coefficient for within-versus between-serotype antigenic change by using 1000 bootstrap samples of the overall (unadjusted) and detrended antigenic time series. For DENV1, DENV2, and DENV4, increasing distance from homotypic strains correlated with increased similarity to heterotypic strains, with the strongest effects for DENV1 and DENV2 and a weaker effect for DENV4 (Fig. 3D). By contrast, DENV3 became more antigenically distant from early homotypic and heterotypic strains linearly and simultaneously (Fig. 3C and D). The largest antigenic change occurred during the replacement of DENV3 genotype II by genotype III between 2010 and 2014 (Fig. 1D). Notably, both genotypes were evolving antigenically relative to homotypic and heterotypic strains, but genotype III achieved greater antigenic distance from other serotypes, especially DENV1 (fig. S5).

We hypothesized that these antigenic changes might be associated with epidemic magnitude over the same period. We estimated the Pearson's correlation coefficient for each antigenic time series with the annual serotype-specific incidence of dengue cases treated at QSNICH in Bangkok (Fig. 4 and fig. S6). On average and independently for DENV1, DENV2, and DENV4, large epidemics occurred when strains were more similar antigenically to other serotypes (Fig. 4, A and B) but less similar to earlier strains of the same serotype (Fig. 4C). By contrast, the lowest DENV3 incidence occurred when DENV3 was most antigenically similar to other DENV3 strains and other serotypes. The aforementioned genotype replacement event followed this period of low incidence, with both DENV3 genotypes evolving antigenically away from earlier DENV3 and heterotypic serotypes as DENV3 incidence rebounded (Fig. 4, A to C, vertical plots). Across serotypes, antigenic change correlated over the entire period and on a year

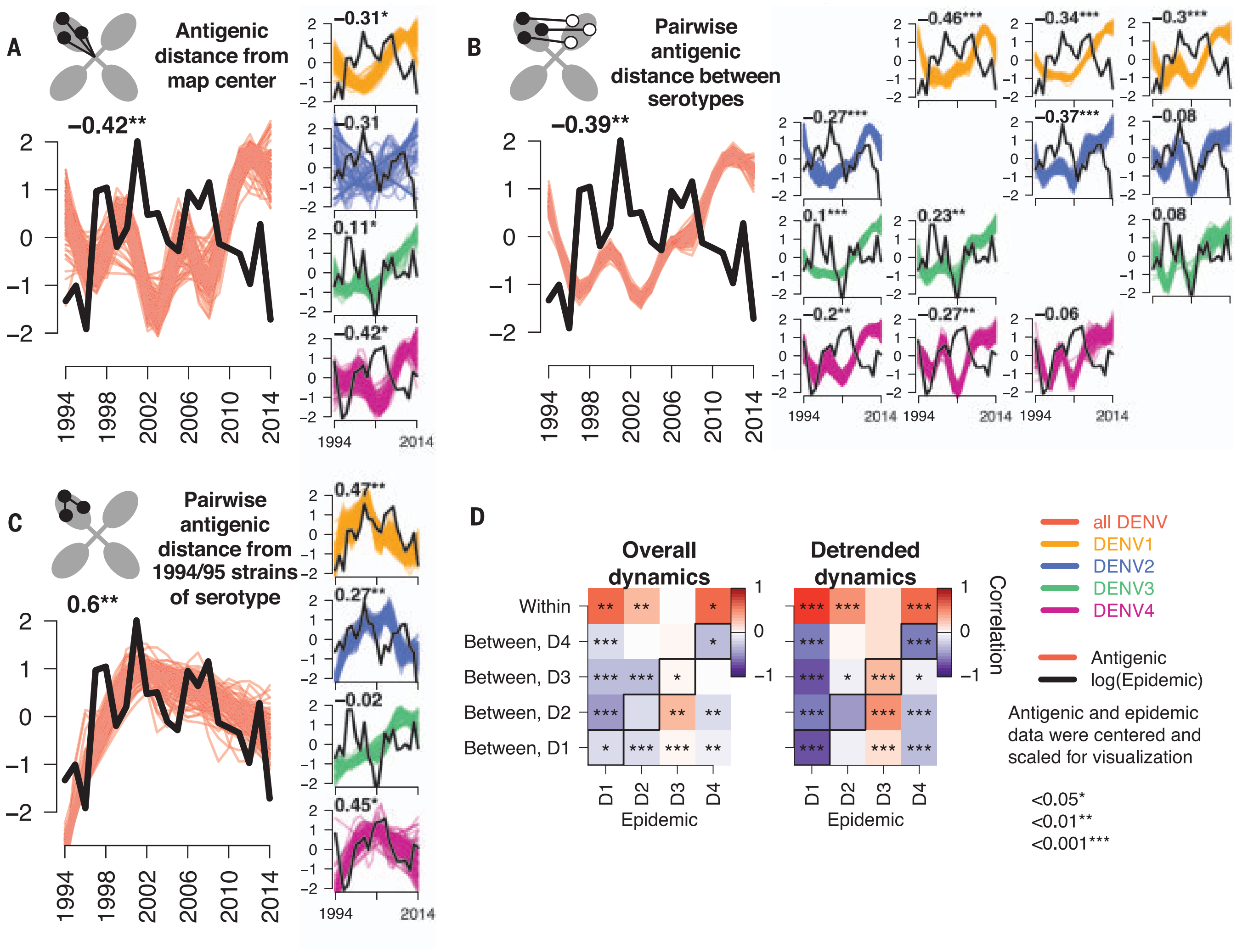


**Fig. 4. Correlation of the antigenic and epidemic time series for DENV1-4 in Bangkok, Thailand, from 1994 to 2014.** Epidemic (black lines) versus antigenic time series (colored lines) for all serotypes (large plots) or each serotype separately (vertical plots) measured as (**A**) distance from the map center, (**B**) pairwise distance between serotypes each year, (**C**) pairwise distance from the 1994 and 1995 strains of the same serotype. All antigenic and epidemic time series are scaled by the standard deviation and centered at zero for visualization. Pearson's correlation coefficients and corresponding *P* values for epidemic versus bootstrapped (*n* = 1000) antigenic time series are shown (top left), with significance indicated by stars. (**D**) Summary of Pearson's correlation coefficients of the epidemic and antigenic time series for overall antigenic dynamics (no adjustment) and detrended dynamics (linear model subtracted from the GAM). Shown for each serotype (columns) and metric of antigenic distance (rows). Diagonal black boxes correspond to distance from the map center; other boxes indicate pairwise distance between the indicated serotypes. Color indicates correlation (range −1 to 1) and significance is indicated by stars.

to year basis with incidence, suggesting a close link between annual epidemiologic change and corresponding shifts in antigenic phenotype at the population level (Fig. 4D and fig. S4, detrended series).

Overall, we found that antigenic evolution differed within a genotype versus during a genotype replacement event. Within genotypes for DENV1, DENV2, and DENV4, major outbreaks correlated with evasion of homotypic protection. A large outbreak of a given genotype could lead to a selective sweep that would be followed by reduced antigenic diversity and thus would decrease antigenic distance between serotypes. Alternatively, the pressure to evade homotypic immunity may be so strong as to drive strains in the direction of other serotypes given that there are structural limits on the amino acids permitted within a given genotype. Strains may also tolerate weak cross-neutralization when such changes improve replication or fitness in mosquitoes or human hosts. For example, laboratory-adapted DENV strains have acquired amino acid changes that render them more susceptible to homotypic and heterotypic neutralization but may be advantageous in cell cultures (*9*, *37*). It is also possible that antigenic change toward other serotypes facilitates antibody-dependent enhancement (*2*). Our assay does not directly measure enhancement, but low neutralizing antibody titers are correlated both with peak enhancement titers and with increased severity of illness (*4*, *38*). Because we sampled dengue cases requiring hospitalization, our study overrepresents severe secondary infections, which may be under stronger immunologic pressure from enhancement and have higher viral loads with greater within-host viral diversity (*29*). Our isolates were collected from a pediatric hospital setting and so may not reflect the full diversity of DENV1-4 circulating in Bangkok during this period. Other studies have linked specific infection histories and viral lineages to increased severity. A study in Nicaragua showed that prior infection with DENV1 versus DENV3 differentially modified disease severity during subsequent DENV2 infection with distinct clades (*17*). In another study, an evolutionarily successful DENV2 lineage had an amino acid change that increased sensitivity to heterotypic neutralization but induced higher viremia during secondary infection (*39*). Increased severity has been theorized to reduce transmission; excessive optimization of cross-serotype enhancement may therefore be an evolutionary dead end, as hospitalization reduces transmission opportunities (*40*). However, the large role that asymptomatic and presymptomatic individuals play in transmission may minimize selection because of this mechanism (*41*). In the case of DENV3, antigenic similarity to other serotypes—as well as to earlier DENV3 strains—was associated with the lowest DENV3 incidence, potentially selecting for DENV3 strains that escaped both homotypic and heterotypic immunity. A previous phylogenetic study of DENV in Thailand showed that clade replacements are associated with declining incidence (*16*). Our findings further support this observation and suggest that immunological pressure imposed by cocirculating serotypes—rather than population bottlenecks—helps govern replacement events. Additional data from genotype replacement events for DENV and other viruses are needed to further evaluate this hypothesis.

The intense cocirculation of multiple DENV serotypes and genotypes in a single location with high population density is relatively new in most parts of the world and likely occurred in Thailand only from the beginning of the 20th century. Even in Thailand, DENV1-4 may still be transitioning to an endemic equilibrium, adjusting antigenic distances relative to one another in an ongoing process. Because of the endemicity of DENV1-4, the serotypes in Thailand may interact more intensively, thus explaining why these serotypes were closer together than those in other regions. The antigenic distance between Thailand serotypes increased over time but by a nonlinear path, with fluctuations in antigenic distance within and among serotypes that closely correlated with epidemic magnitude. Given that DENV epidemic dynamics are governed by population immunity, demography, host behavior, vector abundance, and environmental factors, it is possible that the observed antigenic fluctuations track with longer-term epidemiological patterns that cannot be disentangled with only 20 years of data (*42*). However, our study of cocirculating strains, genotypes, and serotypes suggests that multiple selective mechanisms may affect antigenic evolutionary processes simultaneously, including immune evasion, antibody-mediated enhancement, constraints on viral protein structure, introduction of new genotypes, and local dengue incidence. Specifically, the balance between cross-protection and antibody-dependent enhancement has been posited to explain the phylogenetic distance between DENV1-4 and may help explain the more bounded nature of DENV antigenic evolution compared with other well-studied antigenically variable viruses (*2*). Influenza A and B viruses "zig zag" through antigenic space but fundamentally evolve linearly away from previously circulating strains (*1*, *43*). Although other antigenically variable viruses such as enterovirus 71 and GII.4 noroviruses have complex, nonlinear patterns of variation across antigenic space, the available data do not clearly indicate whether these patterns oscillate over evolutionary time (*44*, *45*). This work constitutes the most comprehensive dataset to date that has been used to explore hypothesized evolutionary tradeoffs for DENV and more broadly among antigenically interacting serotypes, with potentially valuable insights for identifying the determinants of viral antigenic evolution and informing virus surveillance and vaccine evaluation.

## REFERENCES AND NOTES

1. D. J. Smith *et al*., *Science* **305**, 371–376 (2004).
2. B. T. Grenfell *et al*., *Science* **303**, 327–332 (2004).
3. L. C. Katzelnick *et al*., *Science* **358**, 929–932 (2017).
4. H. Salje *et al*., *Nature* **557**, 719–723 (2018).
5. S. B. Halstead, E. J. O'Rourke, *Nature* **265**, 739–741 (1977).
6. S. B. Halstead, *J. Infect. Dis.* **140**, 527–533 (1979).
7. L. Cattarino, I. Rodriguez-Barraquer, N. Imai, D. A. T. Cummings, N. M. Ferguson, Mapping global variation in dengue transmission intensity. *Sci. Transl. Med.* **12**, eaax4144 (2020).
8. S. S. Twiddy, E. C. Holmes, A. Rambaut, *Mol. Biol. Evol.* **20**, 122–129 (2003).
9. L. C. Katzelnick *et al*., *Science* **349**, 1338–1343 (2015).
10. W. B. Messer *et al*., *PLOS Negl. Trop. Dis.* **6**, e1486 (2012).
11. E. N. Gallichotte *et al*., *Cell Rep.* **25**, 1214–1224 (2018).
12. D. R. Martinez *et al*., *Cell Rep.* **33**, 108226 (2020).
13. J. J. Waggoner *et al*., *J. Infect. Dis.* **214**, 986–993 (2016).
14. M. Juraska *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8378–E8387 (2018).
15. S. M. Bell, L. Katzelnick, T. Bedford, *eLife* **8**, 1–22 (2019).
16. C. Zhang *et al*., *J. Virol.* **79**, 15123–15130 (2005).
17. M. OhAinle, *Sci. Transl. Med.* **3**, 114ra128 (2011).
18. B. Adams *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14234–14239 (2006).
19. T. J. Kochel *et al*., *Lancet* **360**, 310–312 (2002).
20. B. M. Forshey *et al*., *PLOS Negl. Trop. Dis.* **10**, e0004398 (2016).
21. Y. Nagao, K. Koelle, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2238–2243 (2008).
22. H. J. Wearing, P. Rohani, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11802–11807 (2006).
23. N. G. Reich *et al*., *J. R. Soc. Interface* **10**, 20130414 (2013).
24. J. Lourenço, M. Recker, *PLOS Comput. Biol.* **9**, e1003308 (2013).
25. A. Ferguson, S. Gupta, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 790–794 (1999).
26. D. A. T. Cummings, I. B. Schwartz, L. Billings, L. B. Shaw, D. S. Burke, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15259–15264 (2005).
27. B. Adams, M. Boots, *J. Theor. Biol.* **242**, 337–346 (2006).
28. H. Salje *et al*., *Science* **355**, 1302–1306 (2017).
29. A. Nisalak *et al*., *Am. J. Trop. Med. Hyg.* **94**, 1342–1347 (2016).
30. D. A. Cummings *et al*., *Nature* **427**, 344–347 (2004).
31. S. B. Halstead, J. E. Scanlon, P. Umpaivit, S. Udomsakdi, *Am. J. Trop. Med. Hyg.* **18**, 997–1021 (1969).
32. E. C. Holmes, S. S. Twiddy, *Infect. Genet. Evol.* **3**, 19–28 (2003).
33. L. C. Katzelnick, *Zenodo* (2021).
34. Z. Moodie *et al*., *J. Infect. Dis.* **217**, 742–753 (2018).
35. L. C. Katzelnick *et al*., *PLOS Negl. Trop. Dis.* **12**, e0006862 (2018).
36. R. de Alwis *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7439–7444 (2012).
37. K. A. Dowd, C. R. DeMaso, T. C. Pierson, Genotypic Differences in Dengue Virus Neutralization Are Explained by a Single Amino Acid Mutation That Modulates Virus Breathing. *mBio* **6**, e01559–e15 (2015).
38. S. C. Kliks, A. Nisalak, W. E. Brandt, L. Wahl, D. S. Burke, *Am. J. Trop. Med. Hyg.* **40**, 444–451 (1989).
39. Wang *et al*., *J. Infect. Dis.* **213**, 975–984 (2016).
40. F. Fenner, I. D. Marshall, *J. Hyg. (Lond.)* **55**, 149–191 (1957).
41. Q. A. ten Bosch *et al*., *PLOS Pathog.* **14**, 82–86 (2018).
42. D. A. Cummings *et al*., *PLOS Med.* **6**, e1000139 (2009).
43. T. Bedford *et al*., *eLife* **3**, e01914 (2014).
44. S. W. Huang *et al*., *J. Clin. Microbiol.* **47**, 3653–3662 (2009).
45. J. A. Kendra, K. Tohma, L. A. Ford-Siltz, C. J. Lepore, G. I. Parra, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015874118 (2021).

## ACKNOWLEDGMENTS

**Funding:** This research was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases (S.S.W. and L.C.K.), National Institute of Allergy and Infectious Diseases and National Institutes of Health Grant R01AI114703-01 (D.A.T.C., L.C.K., A.C.E., A.H., B.G.C., N.C., I.M.B., C.C., L.M., D.S., G.G., R.J., and H.S.), the Military Infectious Disease Research Program (A.H., I.M.B., L.M., G.G., and R.J.), and a European Research Council Grant 804744 (H.S.). Sequencing for infectious disease surveillance was additionally supported by the Global Emerging Infections Surveillance (GEIS) Branch (R.J.). The antigenic cartography toolkit was in part supported by NIAID-NIH Centers of Excellence for Influenza Research and Surveillance contract HHSN272201400008C (D.J.S.) for use on influenza virus. **Author contributions:** L.C.K., D.A.T.C., and H.S. conceived the study, developed and performed analyses, and wrote the manuscript; S.S.W., A.C.E., N.C., C.C., I.M.B., P.B., V.D., P.D., G.G., L.M., B.T., S.F., and R.J. collected the data; D.J.S, A.T.H., and B.G.C. aided in interpretation of the results. All authors commented on and edited the paper. **Competing interests:** P.B. is with GlaxoSmithKline (GSK) vaccines in Singapore and has stock options with GSK. L.M. works at Merck. Material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The views expressed are those of the authors and do not necessarily reflect the official views of the Uniformed Services University of the Health Sciences, the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., the Department of Health and Human Services, the National Institutes of Health, the Departments of the Army, Navy, or Air Force, the Department of Defense, or the U.S. government. **Data and materials availability:** The viruses and antisera used in this study are covered by material transfer agreements between the institutions in this research (WRAIR, NIH, and UF). Requests for sharing of sera or viruses can be directed to the corresponding authors and accommodated subject to institutional and regulatory approvals. R code and raw antigenic and epidemic data used in the figures in this manuscript are available at Zenodo (*33*). All sequence data is publicly available on GenBank (accession numbers KY586306 to KY586946, MW881266, MW945425 to MW945427, MW945430, MW945433 to MW945437, MW945454 to MW945763, MW945772 to MW946604, and MW946607 to MW946985).

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk0058
Materials and Methods
Figs. S1 to S6
Tables S1 to S3
Movie S1
Data S1
References (*46*–*50*)
MDAR Reproducibility Checklist

15 June 2021; accepted 30 September 2021
10.1126/science.abk0058

NATURAL PRODUCTS

# Biosynthesis of fluopsin C, a copper-containing antibiotic from *Pseudomonas aeruginosa*

Jon B. Patteson[1], Andrew T. Putz[1], Lizhi Tao[2], William C. Simke[1], L. Henry Bryant III[1], R. David Britt[2], Bo Li[1,3]*

**Metal-binding natural products contribute to metal acquisition and bacterial virulence, but their roles in metal stress response are underexplored. We show that a five-enzyme pathway in *Pseudomonas aeruginosa* synthesizes a small-molecule copper complex, fluopsin C, in response to elevated copper concentrations. Fluopsin C is a broad-spectrum antibiotic that contains a copper ion chelated by two minimal thiohydroxamates. Biosynthesis of the thiohydroxamate begins with cysteine and requires two lyases, two iron-dependent enzymes, and a methyltransferase. The iron-dependent enzymes remove the carboxyl group and the α carbon from cysteine through decarboxylation, N-hydroxylation, and methylene excision. Conservation of the pathway in *P. aeruginosa* and other bacteria suggests a common role for fluopsin C in the copper stress response, which involves fusing copper into an antibiotic against other microbes.**

Metal-binding natural products allow microorganisms to acquire metal ions for survival and virulence (*1*, *2*). Many microbes produce siderophores for iron acquisition (*1*, *3*, *4*), but few make natural products that bind copper under physiological conditions (*5*). These copper-binding compounds facilitate copper import when copper is limited or help detoxify copper when copper is in excess. Three examples are methanobactins, yersiniabactin, and xanthocillin. Methanobactins are produced by methane-oxidizing bacteria to scavenge copper for copper-dependent methane oxygenase (*6*, *7*). Yersiniabactin, which is a virulence factor produced by uropathogenic *Escherichia coli*, serves a dual role in acquiring copper nutrients and protecting the bacterium from host-derived copper toxicity (*8*, *9*). Xanthocillin and derivatives, which are fungal natural products, help regulate copper levels in the producing fungal strain and inhibit the growth of other fungal and bacterial species (*10*). In the immune system, copper is a bactericide that is marshalled by immune cells to poison invading microbes (*11*, *12*). Thus, copper resistance and detoxification are essential for microbial virulence (*13*, *14*). Although many proteins that help microbes traffic or efflux copper have been uncovered (*15*, *16*), the potential for natural products to participate in copper homeostasis remains mostly unexplored.

We discovered a conserved biosynthetic pathway in *Pseudomonas aeruginosa*, which integrates copper into a broad-spectrum antibiotic in response to elevated copper concentrations. *P. aeruginosa* is a ubiquitous environmental bacterium found in soil, in water, and on human skin. It is responsible for many dangerous infections in hospitalized patients and people with compromised immune systems (*17*). *P. aeruginosa* tolerates high levels of copper (*18*), which is a common and persistent environmental pollutant; given that this bacterium is a prolific producer of natural products, we reasoned that it might synthesize copper-binding compounds to resist excess copper. Transcriptional analysis of *P. aeruginosa* strain PAO1 identified the expression of the operon PA3515–PA3519 as among the most highly (47- to 231-fold) up-regulated genes under high levels of copper stress (*19*). In response to an elevated but nontoxic level of copper, *P. aeruginosa* PAO1 induces expression of PA3515–PA3519 within 5 min, along with *copZ1*, which encodes a copper chaperone, and PA3521–PA3523, which encodes an efflux pump (*20*). This cluster of genes is controlled by the copper-response regulator CueR (Fig. 1A) (*21*), which senses cytoplasmic copper and activates copper efflux genes (*22*). The operon PA3515–PA3519 encodes five proteins of unknown functions, including a putative methyltransferase, two putative adenylosuccinate lyases, and two enzymes that contain a heme oxygenase domain (Fig. 1A). We proposed that these proteins synthesize a copper-binding natural product.

To test this hypothesis and identify the natural product made by the PA3515–PA3519 operon, we obtained mutants in which each gene in the operon is disrupted by a *lacZ* or *phoA* insertion (*23*). We cultured *P. aeruginosa* PAO1 in the presence of 250 μM of $CuSO_4$, a concentration that is well tolerated by *P. aeruginosa*, and compared the metabolomes of wild-type PAO1 with those of individual insertional mutants by means of liquid chromatography coupled–high resolution mass spectrometry (LC-HRMS). We identified two coeluting ions produced only by the wild type with mass-to-charge (*m*/*z*) ratios of 152.929 and 243.938 (Fig. 1B and fig. S1A). Both ions exhibit an isotopic distribution pattern indicative of a copper ion with molecular formulae of $CuC_2H_4NOS^+$ and $CuC_4H_9N_2O_2S_2^+$, respectively (fig. S1A), which suggests that the 243.938 ion is a copper(II) complex that contains two identical ligands. We identified this complex as fluopsin C (Fig. 1C). The complex exhibits a signature d-d transition ultraviolet–visible light (UV-vis) absorption band for copper(II) at 550 to 600 nm (fig. S1B). Fluopsin C is a potent copper-containing antibiotic and cytotoxic agent that was first isolated >50 years ago from environmental *Pseudomonas* and *Streptomyces* strains (*24*–*26*). These early studies revealed that fluopsin C contains two identical *N*-methylthiohydroxamate ligands (Fig. 1C) by means of nuclear magnetic resonance (NMR) and infrared analysis after removal of copper from fluopsin C (*27*, *28*) because the paramagnetic effect of copper(II) prevents direct NMR analysis of the copper complex. Fluopsin C was also identified from the *P. aeruginosa* LV strain that was isolated from a plant lesion, and the compound inhibits the growth of multidrug-resistant bacterial pathogens (*29*). We confirmed that fluopsin C inhibits the growth of a broad spectrum of Gram-positive and Gram-negative bacteria and yeast (table S1). We also found that fluopsin C is effective against multidrug-resistant strains of *Acinetobacter baumannii* and *Staphylococcus aureus* (table S1). By contrast, *P. aeruginosa* strains are much more resistant to fluopsin C than any other bacterial or fungal strains that were tested (*24*, *26*, *28*). For example, *P. aeruginosa* PAO1, the fluopsin C producer, can resist >70 μg/ml of fluopsin C (table S1). Because the structure of fluopsin C is reminiscent of copper pyrithione (fig. S1C), a synthetic copper-thiohydroxamate complex and an important antimicrobial and antifouling agent (*30*), we synthesized copper pyrithione (fig. S2) and measured its efficacy against the same panel of microbes that we also used to test the efficacy of fluopsin C. Except for *P. aeruginosa* PAO1, all microbes that we tested are susceptible to copper pyrithione (table S1 and fig. S2). Thus, PAO1 is resistant to both the fluopsin C it produces and copper pyrithione.

We characterized the structure of the copper-containing fluopsin C by electron paramagnetic resonance (EPR) spectroscopy. EPR confirms the presence of copper(II) ($3d^9$, electron spin $S = 1/2$) in a $d_{x^2-y^2}$ ground state. Toluene was chosen as a solvent because it allows fluopsin C to produce well-defined EPR signals at cryogenic temperatures. When fluopsin C is dissolved in toluene, two components are observed in ~1:1 ratio in the solution state at room temperature (Fig. 1D), whereas one component dominates (85%) at cryogenic temperature (15 K) in the frozen-solution state (Fig. 1, E and F). Although the EPR signals of these components overlap through the entire X-band envelope at the microwave frequency of 9.39 GHz (Fig. 1E), the signals are well separated at the Q-band (34.0 GHz), at which only the major component

---

[1] Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [2] Department of Chemistry, University of California, Davis, Davis, CA, USA. [3] Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
*Corresponding author. Email: boli@email.unc.edu

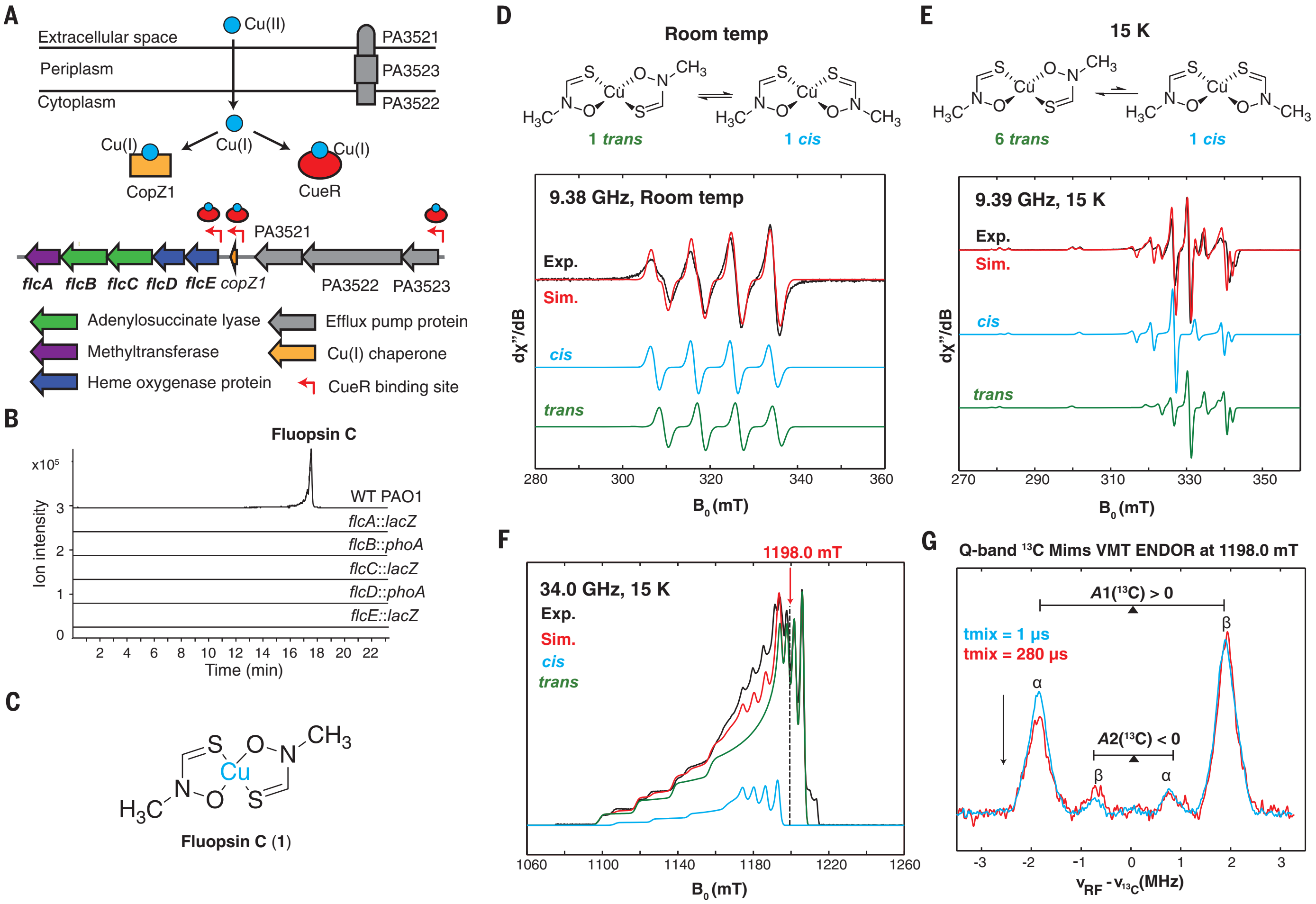


**Fig. 1. Identification of the fluopsin C gene cluster (*flc*).** (**A**) The *flc* cluster (PA3515–PA3519) in *P. aeruginosa* PAO1 is located in a copper island regulated by CueR. PA3521–PA3523 encodes an efflux pump (MexPQ-OpmE). (**B**) Genes *flcA* to *flcE* in PAO1 are essential for fluopsin C production in $CuSO_4$-containing media. Extracted ion chromatograms (EICs) of fluopsin C (**1**) (*m/z* 152.9304 $[M-C_2H_4NOS]^+$) are shown. (**C**) Structure of fluopsin C. (**D**) X-band room temperature EPR spectra of fluopsin C in toluene show the cis isomer ($g_{iso}$ = 2.0900, $A_{iso}(^{63}Cu)$ = 259 MHz) and trans isomer ($g_{iso}$ = 2.0808, $A_{iso}(^{63}Cu)$ = 247 MHz) with approximately equal concentrations. $g_{iso}$, isotropic *g* value; $A_{iso}$, isotropic *A* value; Exp., experimental; Sim., simulated. (**E**) X-band frozen-solution EPR spectra of fluopsin C in toluene at 15 K show different electronic structures for the cis isomer (*g* = [2.151, 2.064, 2.064], $A(^{63}Cu)$ = [580, 173, 173] MHz) and trans isomer (*g* = [2.166, 2.036, 2.036], $A(^{63}Cu)$ = [583, 108, 108] MHz) in ~1:6 ratio. (**F**) Q-band electron spin-echo–detected field-swept EPR spectrum confirms 1:6 ratio of cis:trans. (**G**) Q-band variable mixing time (VMT) Mims-ENDOR spectra of $[^{15}N,^{13}C_2]$-fluopsin C in toluene at the magnetic field position (1198.0 mT) with contribution from only the trans isomer. tmix, mixing time.

contributes to the EPR envelope at magnetic field positions higher than 1198.0 mT (Fig. 1F). The electron-nuclear double resonance (ENDOR) spectra of $[^{15}N,^{13}C_2]$-fluopsin C at 1198.0 mT (Fig. 1G) exhibit two $^{13}C$ hyperfine coupling signals with different magnitudes and opposite signs ($A1 \approx +4$ MHz, $A2 \approx -2$ MHz). Combining the ENDOR data with theoretical calculations (fig. S3), we identify the major component as the trans isomer and assign the hyperfine coupling values *A*1 and *A*2 to the methyl and thiohydroxamate $^{13}C$, respectively. Thus, the trans isomer of fluopsin C is favored (cis:trans = 1:6) in the frozen-solution state (Fig. 1, E and F).

We also assessed the stability of fluopsin C in the presence of Fe(III) and reductants in aqueous solutions. Fluopsin C remains intact in the presence of Fe(III) at pH 7.4, as evidenced by minimal changes to the UV-vis spectrum of fluopsin C (fig. S1D). Titrating 1 to 5 equiv of dithiothreitol or glutathione into fluopsin C resulted in a notable shift in the UV-vis spectrum, including a reduction of the d-d absorption band for Cu(II) at 550 to 600 nm (fig. S1, E and F), which is evidence of the reduction of Cu(II) to Cu(I) in fluopsin C. These results suggest that fluopsin C exists as a Cu(I) complex in the reducing cytoplasm.

The structure of fluopsin C bears little resemblance to that of other metal-binding natural products. Specifically, the thiohydroxamate differs from other copper ligands, including the paired thioamides and oxazolone heterocycles in methanobactins (*6*), the thiazoli(di)nes in yersiniabactin (*31*), and the diisonitriles in xanthocillin (*10*), as well as the diisonitriles in other compounds known to mediate copper or zinc acquisition (*32*, *33*). A natural thiohydroximate with *S*-glucosyl-*O*-sulfonyl moieties is found in the plant defense compounds glucosinolates (fig. S1C) (*34*). In comparison, the thiohydroxamate in fluopsin C is minimal, containing only five nonhydrogen atoms. Although the structure and antimicrobial activity of fluopsin C have been known since 1970, its biosynthesis and biological function were unknown.

To confirm the role of PA3515–PA3519 in fluopsin C biosynthesis, we expressed this operon under a *lacUV5* promoter in a heterologous host, *Pseudomonas fluorescens* SBW25, which lacks homologous genes. *P. fluorescens* that expresses the operon produces fluopsin C in the presence of $CuSO_4$ (fig. S4), which indicates that PA3515–PA3519 is sufficient for fluopsin C production and confirms our assignment of the biosynthetic gene cluster (*flc*, PA3515–PA3519 as *flcA* to *flcE*). Having identified the *flc* cluster, we investigated the biosynthetic pathway that generates the unusual structure of fluopsin C. We cultured *P. aeruginosa* PAO1 in M9 minimal media that contained $CuSO_4$ (fig. S5) and conducted feeding studies by using either $[^{15}N,^{13}C_5]$ L-methionine or $[^{15}N,^{13}C_3]$ L-cysteine as potential precursors. Analysis by LC-HRMS revealed that one carbon from $[^{15}N,^{13}C_5]$ L-methionine and one nitrogen and one carbon from $[^{15}N,^{13}C_3]$ L-cysteine were incorporated in each ligand of fluopsin C (Fig. 2A). We reasoned that the *N*-methyl comes from methionine, likely through an *S*-adenosyl-L-methionine (SAM)–dependent methylation, and the thiohydroxamate backbone comes from L-cysteine.

To identify the first biosynthetic step for fluopsin C, we overexpressed each Flc protein in *E. coli* and analyzed the metabolomes of the corresponding strains by LC-HRMS. The FlcB-overexpressing *E. coli* produced a species with a $m/z$ of 238.040 (**2**, $C_7H_{12}NO_6S^+$) at a substantially higher level than the other strains (fig. S6). This mass is consistent with *S*-succinyl cysteine. FlcB belongs to the adenylosuccinate lyase superfamily that catalyzes the cleavage of adenylosuccinate into adenine monophosphate and fumarate (*35*). Thus, we hypothesized that FlcB catalyzes a reverse reaction: conjugate addition of L-cysteine to fumarate to form **2**. Indeed, purified FlcB catalyzed the in vitro conversion of L-cysteine and fumarate to a single isomer of **2** with an apparent turnover number ($k_{cat}$) of 0.74/s, whereas a small amount of racemic **2** slowly formed in the absence of FlcB (Fig. 2B and fig. S7), indicating that FlcB is required for efficient and stereoselective synthesis of **2**. NMR characterization reveals identical chemical shifts between **2** and the synthetic *R* isomer but not the *S* isomer (figs. S8 to S13 and table S3). Thus, FlcB catalyzes the first step in fluopsin C biosynthesis by conjugating L-cysteine with fumarate (Fig. 2B).

We investigated the remaining Flc enzymes for activity toward **2**. Each purified enzyme (FlcA, FlcC, FlcD, or FlcE) was incubated with **2** in the presence of *E. coli* cell-free lysate (figs. S14 and S15). The incubation of FlcE with **2** produced a distinct species, **3** ($m/z$ 208.028, $C_6H_{10}NO_5S^+$) (Fig. 2C and fig. S15), which corresponds to the net loss of C, O, and two H from **2**. FlcE can modify (*R*)-**2** but not (*S*)-**2**, which confirms that the first intermediate is (*R*)-**2** (fig. S16). Sequence alignment suggests that both FlcD and FlcE contain a heme oxygenase domain. Heme oxygenase domains also exist in the *Chlamydia* protein associating with death domains (CADD) and the enzyme SznF, and the crystal structures of these proteins show that this domain contains a diiron center (*36*, *37*). Therefore, we incubated FlcE with **2** in the presence of Fe(II) without cell lysate and observed the formation of **3** (Fig. 2C). The activity of FlcE requires Fe(II) and is only modestly increased by addition of ascorbate as a reductant; no other common metal cofactors reconstitute FlcE activity (Fig. 2C and fig. S17). These results suggest that FlcE is an iron-dependent enzyme that does not require an external reductant. NMR characterization of **3** indicates that it lacks the cysteine amine or carboxylate but contains an oxime instead (Fig. 2D, tables S4 and S5, and figs. S18 to S22),

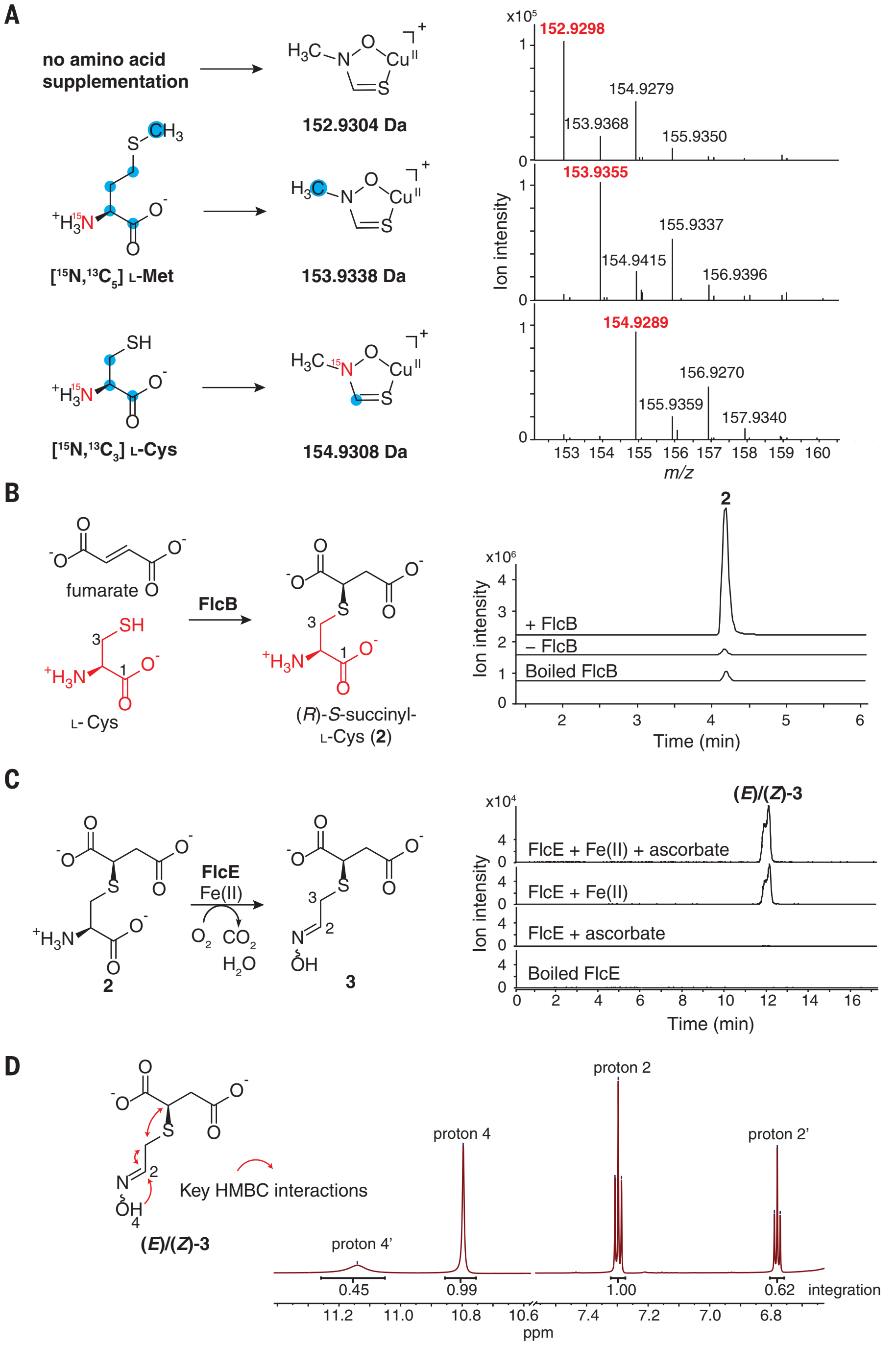


**Fig. 2. FlcB catalyzes S-succinylation of cysteine, and FlcE performs oxidative decarboxylation.** (**A**) Feeding studies in *P. aeruginosa* PAO1 using $[^{15}N,^{13}C_5]$ L-methionine or $[^{15}N,^{13}C_3]$ L-cysteine. Mass spectra of fluopsin C with a single ligand are shown. Positions of isotopic labeling are inferred. (**B**) EICs of **2** ($m/z$ 238.0380 $[M+H]^+$) from the FlcB reaction and controls. A small amount of racemic **2** forms nonenzymatically. (**C**) FlcE converts **2** to **3** in an Fe(II)-dependent manner. EICs of **3** ($m/z$ 230.0094 $[M+Na]^+$) are shown. (**D**) NMR spectrum of **3** in deuterated dimethylsulfoxide supports an oxime structure that exists in two diastereomers. HMBC, heteronuclear multiple bond correlation; ppm, parts per million.

which suggests that FlcE catalyzes both an oxidative decarboxylation and N-hydroxylation.

After identifying the reaction catalyzed by FlcE, we found that the other heme oxygenase domain–containing enzyme, FlcD, catalyzes the conversion of **3** to a distinct species, **4** (*m/z* 194.012, $C_5H_8NO_5S^+$) in the presence of Fe(II) (Fig. 3A and fig. S23). The mass difference between **3** and **4** corresponds to a loss of $CH_2$. Only Fe(II), among the common metal cofactors, reconstitutes the activity of *apo*-FlcD, and the reductant ascorbate modestly increases FlcD activity (Fig. 3A and fig. S24). We isolated **4** from a one-pot reaction containing **2**, FlcD, and FlcE for structural characterization. NMR analysis revealed that **4** also contains an oxime and is one methylene shorter than **3** (table S6 and figs. S25 to S29), which indicates that FlcD catalyzes a methylene excision of **3**. To identify which carbon atom is excised by FlcD, we conducted a one-pot reaction containing FlcB, FlcE, and FlcD using [3-$^{13}$C] L-cysteine as a precursor to **4**. The 3-$^{13}$C atom in [3-$^{13}$C] L-cysteine is retained in **4**, which suggests excision of 2-C by FlcD (Fig. 3B). To determine the fate of the excised 2-C, we examined formaldehyde and formate as potential oxidation products formed by FlcD. No formaldehyde was detected (fig. S30), whereas formate and **4** were produced in 1:1 stoichiometry (Fig. 3C and figs. S31 and S32). Therefore, FlcD catalyzes excision of 2-C from **3** as formate in a net four-electron oxidation. The four-electron oxidation and the lack of need for external reductants suggest that FlcD incorporates oxygens from $O_2$ into formate while regenerating the ferrous center (fig. S33).

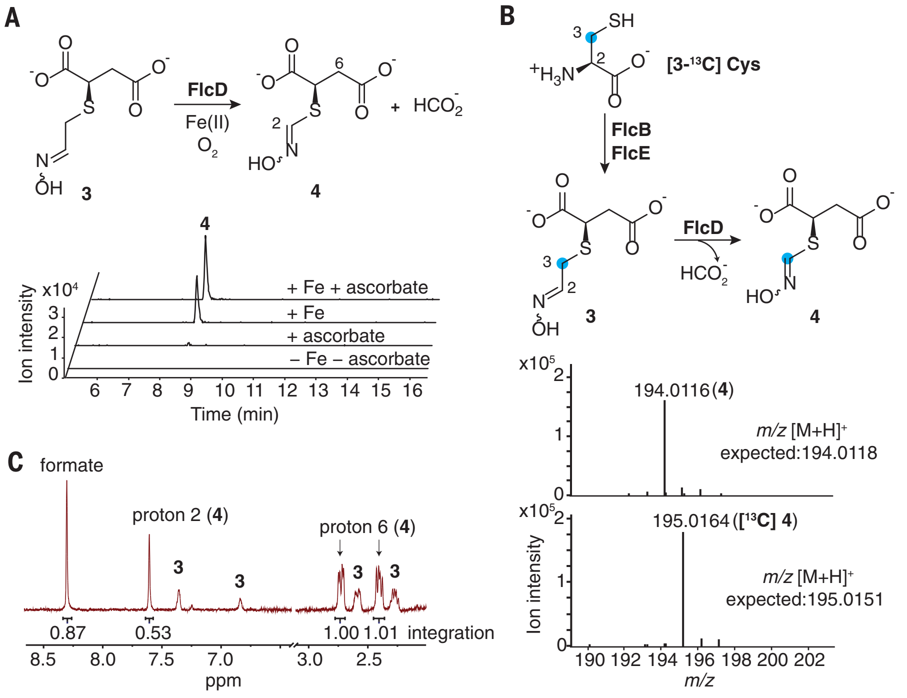


**Fig. 3. FlcD excises a methylene carbon from 3 as formate.** (**A**) FlcD-catalyzed formation of **4** (*m/z* 194.0118 [M+H]$^+$) depends on Fe(II). (**B**) Mass spectra show that the 3-$^{13}$C atom of [3-$^{13}$C] L-cysteine is retained in **4** in a one-pot reaction containing FlcB, FlcE, and FlcD. (**C**) NMR analysis of the FlcD reaction in $D_2O$ reveals the production of formate and **4** in a 1:1 stoichiometry.

Comparing the structure of **4** to that of fluopsin C, we hypothesized that the remaining two enzymes, the putative methyltransferase FlcA and lyase FlcC, catalyze N-methylation and cleavage of thiohydroxamate from succinate, respectively. Purified **4** was incubated with either FlcA or FlcC. FlcA does not catalyze methylation of **4** in the presence of SAM, whereas FlcC rapidly catalyzes consumption of **4** and production of fumarate (figs. S34 and S35). Monitoring the FlcC reaction by NMR reveals a time-dependent consumption of **4** (apparent $k_{cat}$ = 0.46/s) and concomitant production of fumarate and a distinct species, **5** (Fig. 4A and fig. S35). Compound **5** contains a sole slowly exchanged singlet proton at 8.0 parts per million in $D_2O$ (Fig. 4A and fig. S36) and does not react with Ellman's reagent that modifies free thiols (fig. S37), which suggests that **5** is a thiohydroxamate rather than an oxime thiol. Therefore, FlcC catalyzes the C–S bond cleavage of **4** to yield the thiohydroxamate **5** and fumarate.

To reconstitute the activity of the final methyltransferase FlcA, we incubated **4** with FlcC, FlcA, and SAM, adding $CuSO_4$ at the end of the reaction. Fluopsin C is produced only when both Flc enzymes are present (Fig. 4B), supporting the role of FlcA in methylating **5** and generating the Cu-free fluopsin (**6**). Having reconstituted each biosynthetic step, we performed a one-pot reaction containing L-cysteine, fumarate, and all five Flc enzymes and their cofactors, and we added 2 mM $CuSO_4$ at the end. A 24-hour one-pot reaction containing 0.25 to 0.75 mol % of each Flc enzyme resulted in an 87% yield across five steps. Omission of any Flc enzyme abolished the production of fluopsin C; in particular, omission of FlcE, FlcD, or FlcC resulted in the accumulation of **2**, **3**, and **4**, respectively, which further supports the biosynthetic pathway (fig. S38 and Fig. 4C).

We uncovered the biosynthetic origin of fluopsin C, a metal complex produced by bacteria, and characterized its electronic structure. Its biosynthesis entails a series of unusual enzymatic transformations that convert cysteine to a minimal thiohydroxamate (Fig. 4C). The lyases FlcB and FlcC add and subsequently remove an *S*-succinyl group while recycling fumarate. The heme oxygenase enzymes, FlcE and FlcD, remove both 1-C and 2-C from cysteine and oxidize the amine to an oxime, generating the thiohydroximate form that is protected by the *S*-succinyl group. Only a few heme oxygenase enzymes have been characterized, such as SznF in the biosynthesis of the anticancer drug streptozotocin (*38*), UndA in the biosynthesis of the biofuel undecene (*39*, *40*), and BesC in the biosynthesis of a terminal alkyne-containing amino acid (*41*). FlcD and FlcE share low sequence identity with these enzymes and contain divergent sequences at conserved positions (figs. S39 and S40), which may hint at different mechanisms used by FlcD and FlcE. In particular, the carbon excision reaction catalyzed by FlcD has limited precedence in enzyme chemistry (*42*, *43*).

Nonclassical roles of siderophores are emerging, such as the transport of noniron metals (*44*). Our discovery of the fluopsin C biosynthetic pathway further expands the functions of metal-binding natural products beyond metal acquisition. As we and others have shown, fluopsin C exhibits broad-spectrum antibiotic activity against microbes and potent cytotoxicity against mammalian cells and whole animals, and it is much less active against various *P. aeruginosa* strains (*24*, *26*, *29*). Because fluopsin C contains copper, its antimicrobial mechanism likely differs from existing antibiotic classes or the copper-binding compound xanthocillin, which exhibits diminished antimicrobial effect in the presence of copper (*10*). We found that the *flc* gene cluster is preserved in 4937 of the 4955 (99.6%) *P. aeruginosa* genomes, which include both clinical and environmental isolates (*45*). The preservation of the *flc* cluster suggests that almost all *P. aeruginosa* strains can produce and resist fluopsin C. Given that elevated levels of copper occur in water, soil, and activated macrophages (*2*, *46*), *P. aeruginosa*

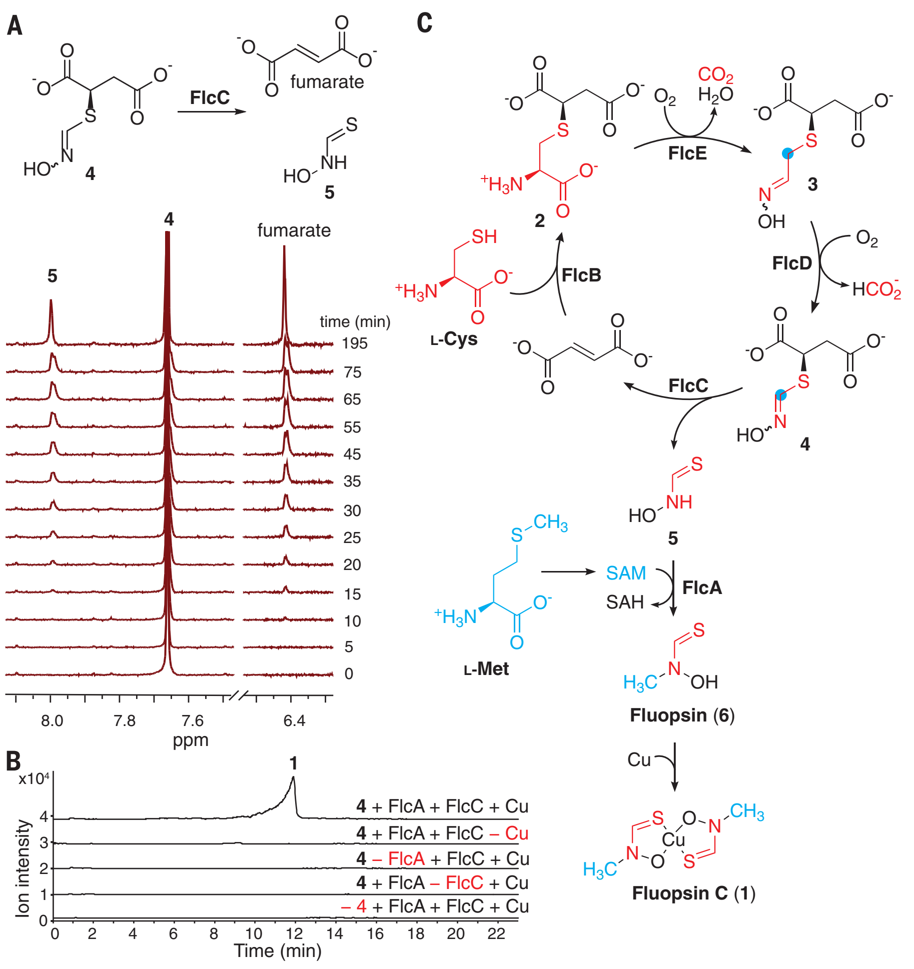


**Fig. 4. FlcC and FlcA catalyze the final steps of fluopsin C biosynthesis.** (**A**) NMR time-course study of the lyase FlcC. (**B**) FlcA and FlcC catalyze tandem transformations of **4** to fluopsin C, with addition of $CuSO_4$ at the end of the reaction. Combined EICs of **1** (*m/z* 152.9304 $[M–C_2H_4NOS]^+$ and 265.9215 $[M+Na]^+$) are shown. (**C**) Complete fluopsin C biosynthetic pathway. SAH, *S*-adenosyl-L-homocysteine.

may integrate copper into fluopsin C to inhibit the growth of other microbes or the activities of host cells during infection as part of the copper stress response. The *flc* cluster is also distributed in the genomes of many other bacteria (figs. S41 and S42 and table S7), including *A. baumannii*, *Enterobacter cloacae*, and *Streptococcus dysgalactiae*, particularly isolates from infected human patients and in hospital settings. Additionally, we identified the *flc* cluster in the genomes of *P. fluorescens*, *Lysobacter enzymogenes*, and *Lysobacter capsica* strains (fig. S41 and S42 and table S7), which colonize the roots and leaves of plants and exert biocontrol activity against phytopathogenic organisms. In particular, the *flc*-containing *L. capsica* AZ78 strain is resistant to copper and more effectively protects plants against oomycetes in combination with a copper fungicide (*47*); the production of fluopsin C may contribute to the copper resistance and copper-augmented fungicidal activity of this strain. Overall, the wide distribution of *flc* in both pathogenic and environmental bacteria suggests that fluopsin C serves a conserved role in helping the bacteria that produce it to interact with other microbes in the environment and their eukaryotic hosts in response to elevated copper levels.

## REFERENCES AND NOTES

1. M. Sandy, A. Butler, *Chem. Rev.* **109**, 4580–4595 (2009).
2. L. D. Palmer, E. P. Skaar, *Annu. Rev. Genet.* **50**, 67–91 (2016).
3. G. L. Challis, *ChemBioChem* **6**, 601–611 (2005).
4. G. Ghssein *et al.*, *Science* **352**, 1105–1109 (2016).
5. G. E. Kenney, A. C. Rosenzweig, *Annu. Rev. Biochem.* **87**, 645–676 (2018).
6. H. J. Kim *et al.*, *Science* **305**, 1612–1615 (2004).
7. M. O. Ross *et al.*, *Science* **364**, 566–570 (2019).
8. K. S. Chaturvedi, C. S. Hung, J. R. Crowley, A. E. Stapleton, J. P. Henderson, *Nat. Chem. Biol.* **8**, 731–736 (2012).
9. E.-I. Koh, A. E. Robinson, N. Bandara, B. E. Rogers, J. P. Henderson, *Nat. Chem. Biol.* **13**, 1016–1021 (2017).
10. N. Raffa *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015224118 (2021).
11. C. White, J. Lee, T. Kambe, K. Fritsche, M. J. Petris, *J. Biol. Chem.* **284**, 33949–33956 (2009).
12. K. Y. Djoko, C. L. Ong, M. J. Walker, A. G. McEwan, *J. Biol. Chem.* **290**, 18954–18961 (2015).
13. F. Wolschendorf *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1621–1626 (2011).
14. J. R. Sheldon, E. P. Skaar, *Curr. Opin. Immunol.* **60**, 1–9 (2019).
15. D. L. Huffman, T. V. O'Halloran, *Annu. Rev. Biochem.* **70**, 677–701 (2001).
16. Y. Fu, F.-M. J. Chang, D. P. Giedroc, *Acc. Chem. Res.* **47**, 3605–3613 (2014).
17. P. D. Lister, D. J. Wolter, N. D. Hanson, *Clin. Microbiol. Rev.* **22**, 582–610 (2009).
18. G. M. Teitzel, M. R. Parsek, *Appl. Environ. Microbiol.* **69**, 2313–2320 (2003).
19. G. M. Teitzel *et al.*, *J. Bacteriol.* **188**, 7242–7256 (2006).
20. J. Quintana, L. Novoa-Aponte, J. M. Argüello, *J. Biol. Chem.* **292**, 15691–15704 (2017).
21. J. T. Thaden, S. Lory, T. S. Gardner, *J. Bacteriol.* **192**, 2557–2568 (2010).
22. A. Changela *et al.*, *Science* **301**, 1383–1387 (2003).
23. K. Held, E. Ramage, M. Jacobs, L. Gallagher, C. Manoil, *J. Bacteriol.* **194**, 6387–6389 (2012).
24. Y. Egawa, K. Umino, S. Awataguchi, Y. Kawano, T. Okuda, *J. Antibiot.* **23**, 267–270 (1970).
25. H. Otsuka *et al.*, *J. Antibiot.* **25**, 369–370 (1972).
26. S. Ito, K. Inuzuka, T. Suzuki, *J. Antibiot.* **23**, 542–545 (1970).
27. Y. Egawa, K. Umino, Y. Ito, T. Okuda, *J. Antibiot.* **24**, 124–130 (1971).
28. K. Shirahata, T. Deguchi, T. Hayashi, I. Matsubara, T. Suzuki, *J. Antibiot.* **23**, 546–550 (1970).
29. M. O. P. Navarro *et al.*, *Front. Microbiol.* **10**, 2431 (2019).
30. M. E. Helsel, E. J. White, S. Z. A. Razvi, B. Alies, K. J. Franz, *Metallomics* **9**, 69–81 (2017).
31. H. Drechsel *et al.*, *Liebigs Ann.* **1995**, 1727–1733 (1995).
32. L. Wang *et al.*, *ACS Chem. Biol.* **12**, 3067–3075 (2017).
33. N. C. Harris *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7025–7030 (2017).
34. N. K. Clay, A. M. Adio, C. Denoux, G. Jander, F. M. Ausubel, *Science* **323**, 95–101 (2009).
35. E. A. Toth, T. O. Yeates, *Structure* **8**, 163–174 (2000).
36. R. Schwarzenbacher *et al.*, *J. Biol. Chem.* **279**, 29320–29324 (2004).
37. M. J. McBride *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015931118 (2021).
38. T. L. Ng, R. Rohac, A. J. Mitchell, A. K. Boal, E. P. Balskus, *Nature* **566**, 94–99 (2019).
39. B. Zhang *et al.*, *J. Am. Chem. Soc.* **141**, 14510–14514 (2019).
40. O. M. Manley, R. Fan, Y. Guo, T. M. Makris, *J. Am. Chem. Soc.* **141**, 8684–8688 (2019).
41. J. A. Marchand *et al.*, *Nature* **567**, 420–424 (2019).
42. R. M. Cicchillo *et al.*, *Nature* **459**, 871–874 (2009).
43. C. P. Ting *et al.*, *Science* **365**, 280–284 (2019).
44. T. C. Johnstone, E. M. Nolan, *Dalton Trans.* **44**, 6320–6339 (2015).
45. G. L. Winsor *et al.*, *Nucleic Acids Res.* **44**, D646–D653 (2016).
46. D. Wagner *et al.*, *J. Immunol.* **174**, 1491–1500 (2005).
47. G. Puopolo, O. Giovannini, I. Pertot, *Microbiol. Res.* **169**, 633–642 (2014).
48. J. B. Patteson *et al.*, Heme oxygenase pfam14518 sequence similarity network, Zenodo (2021); https://doi.org/10.5281/zenodo.5508248.

## ACKNOWLEDGMENTS

We thank W. van der Donk, G. Pielak, A. Bowers, E. Patteson, and members of the Li lab for helpful discussions and K. Trippe (USDA-ARS National Forage Seed Production Research Center) for *P. fluorescens* SBW25. **Funding:** This work was supported by National Institutes of Health grant DP2HD094657 (B.L.), the Packard Fellowship for Science and Engineering (B.L.), The University of North Carolina at Chapel Hill (B.L.), and National Institutes of Health grant R35GM126961 (R.D.B.). NMR experiments were performed in the University of North Carolina at Chapel Hill Department of Chemistry NMR Core Laboratory supported in part by National Science Foundation grants CHE-0922858 and CHE-1828183. **Author contributions:** Conceptualization: J.B.P. and B.L. Methodology: J.B.P., A.T.P., L.T., and W.C.S. Investigation: J.B.P., A.T.P., L.T., W.C.S., and L.H.B. Funding acquisition: B.L. and R.D.B. Project administration: B.L. Supervision: B.L. and R.D.B. Writing – original draft: J.B.P. and B.L. Writing – review and editing: J.B.P., B.L., A.T.P., W.C.S., L.T., and R.D.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The sequence similarity network is available on Zenodo (*48*). Materials are available by request from the corresponding author.

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj6749
Materials and Methods
Figs. S1 to S42
Tables S1 to S7
References (*49–77*)
MDAR Reproducibility Checklist

2 June 2021; accepted 1 October 2021
10.1126/science.abj6749

NEUROSCIENCE

# Fear balance is maintained by bodily feedback to the insular cortex in mice

**Alexandra S. Klein[1,2], Nate Dolensek[1,3], Caroline Weiand[1,4]†, Nadine Gogolla[1]*†**

**How does the brain maintain fear within an adaptive range? We found that the insular cortex acts as a state-dependent regulator of fear that is necessary to establish an equilibrium between the extinction and maintenance of fear memories in mice. Whereas insular cortex responsiveness to fear-evoking cues increased with their certainty to predict harm, this activity was attenuated through negative bodily feedback that arose from heart rate decelerations during freezing. Perturbation of body-brain communication by vagus nerve stimulation disrupted the balance between fear extinction and maintenance similar to insular cortex inhibition. Our data reveal that the insular cortex integrates predictive sensory and interoceptive signals to provide graded and bidirectional teaching signals that gate fear extinction and illustrate how bodily feedback signals are used to maintain fear within a functional equilibrium.**

Fear is essential for survival but has to be maintained within a functional balance that minimizes risk-taking while allowing for the pursuit of important needs (*1*). Next to defensive behaviors, fear evokes strong bodily reactions, such as changes in heart and breathing rates (*2*–*5*). Although bodily feedback signals to the brain are thought to play a crucial role in emotion regulation (*6*–*9*), the neural mechanisms that mediate fear regulation through bodily feedback are incompletely understood. The insular cortex (InsCtx) is a core region involved in the processing of bodily signals, also referred to as interoception (*10*). It receives strong inputs from distinct thalamic and brainstem nuclei that transmit visceral and cardiovascular signals from the periphery to the brain (*11*–*13*). Data from human and other animals have implicated the InsCtx in fear and extinction learning, as well as the learning of safety (*14*–*21*). In addition, human neuroimaging studies have identified structural and functional alterations of the InsCtx as a hallmark of anxiety disorders (*6*, *22*). We therefore sought to address how the interoceptive InsCtx may contribute to fear regulation.

To investigate the influence of InsCtx activity on fear regulation, we performed classical auditory fear conditioning (FC) and used optogenetic inhibition during fear extinction learning in mice (Fig. 1A). Viral vectors leading to excitatory-neuron specific expression of an inhibitory opsin [halorhodopsin (NpHR)] or a control protein [enhanced yellow fluorescent protein (eYFP)] were bilaterally injected into the visceral InsCtx. Optical fibers were placed above the injection sites (Fig. 1B and fig. S1). Efficacy of NpHR-mediated neuronal inhibition in behaving animals was confirmed using extracellular recordings of single units (fig. S2). We first used a relatively weak FC protocol with mildly aversive footshocks as unconditioned stimuli (US), resulting in moderate fear levels [measured as freezing in response to the conditioned stimulus (CS+) during recall; see supplementary materials for details on experimental design and a glossary of key terms]. InsCtx inhibition during the presentation of the CS+ throughout the extinction session led to acute facilitation of extinction and improved extinction performance (measured as the normalized freezing decrease between recall and late extinction; Fig. 1C and fig. S3, A and B). Human studies suggest that InsCtx engagement as well as the intensity of bodily reactions increase proportional to fear levels (*3*, *6*, *23*–*26*). We therefore next assessed the dependence of InsCtx activity manipulations on the level of fear. We used a strong FC protocol by enhancing the US, which produced higher fear levels in reaction to the CS+. Paradoxically, inhibiting InsCtx during extinction after strong FC led to the opposite results as observed upon weak FC, namely an impairment in extinction learning (Fig. 1D and fig. S3, A and C).

These results raised the question of whether the observed opposite effects upon the same activity manipulation were caused by different fear levels resulting from differently strong FC protocols, individual differences in fear reactivity to an aversive experience, or a combination of both. We therefore designed a consecutive FC paradigm, where the same mice subsequently underwent first weak and then strong FC (Fig. 1, E to H; fig. S4; and materials and methods). This experiment confirmed the initial observations (Fig. 1E, left) that InsCtx inhibition upon different FC protocols produced opposite effects. Whereas control mice exhibited stable fear extinction performances after both FC sessions, upon InsCtx inhibition, extinction performances were enhanced after weak FC but reduced after strong FC (Fig. 1E, right). To assess the influence of individual differences in fear reactivity, we assessed each animal's fear level during recall after the first FC session (measured as freezing in response to the CS+) and correlated it to their extinction performance after both weak and strong FC. Whereas in control animals individual fear levels did not affect extinction performances, upon InsCtx inhibition, fear extinction performances were dependent on individual fear levels. Animals that exhibited high fear levels during the recall session extinguished less efficiently, and animals that exhibited low fear levels during the recall session extinguished more rapidly (Fig. 1F and fig. S4C).

To address how individual differences in reactivity to an aversive experience and conditioning strength interacted, we divided animals into low- and high-fear groups based on their individual fear level and assessed their extinction performance separately for strong and weak FC. Again, control animals exhibited stable extinction performances across different combinations of individual fear levels and strengths of aversive experiences (Fig. 1G). However, InsCtx inhibition affected extinction performances only in situations where fear level and experience were on extreme ends: high-fear animals experiencing strong FC or low-fear animals experiencing weak FC (Fig. 1, G and H; and fig. S4D). Intermediate combinations of aversive experience and fear level (weak FC in high-fear animals or strong FC in low-fear animals) were unaffected by InsCtx inhibition (Fig. 1G).

Our results so far revealed that the InsCtx gates extinction learning acutely. We next tested whether InsCtx inhibition also had persistent effects. To avoid complete extinction of fear as observed upon prolonged extinction training (see figs. S3 and S4), mice underwent a shortened extinction session under InsCtx inhibition and were tested on a subsequent day for extinction memories during retrieval in absence of InsCtx inhibition. Like the acute effects during extinction, extinction memories during retrieval were largely stable across different fear levels in control animals, whereas InsCtx inhibition rendered extinction retrieval depending on initial fear levels (Fig. 1, I and J).

To assess how the InsCtx may gate fear extinction, we recorded the responsiveness of the InsCtx to fear-associated cues across fear conditioning and extinction by fiber photometry (*27*) (Fig. 2A). We first quantified InsCtx responses averaged across all animals. Whereas during habituation the presentation of the initially innocuous auditory stimuli did not activate the InsCtx, the InsCtx was increasingly activated by successive CS+ and US pairings during FC. CS+ evoked InsCtx responses

---

[1]Circuits for Emotion Research Group, Max Planck Institute of Neurobiology, Martinsried, Germany. [2]International Max-Planck Research School for Molecular Life Sciences, Munich, Germany. [3]Graduate School of Systemic Neurosciences, Ludwig-Maximilians University, Munich, Germany. [4]International Max-Planck Research School for Translational Psychiatry, Munich, Germany.
***Corresponding author. Email: ngogolla@neuro.mpg.de**
†Present address: Emotion Research Department, Max Planck Institute of Psychiatry, Munich, Germany.

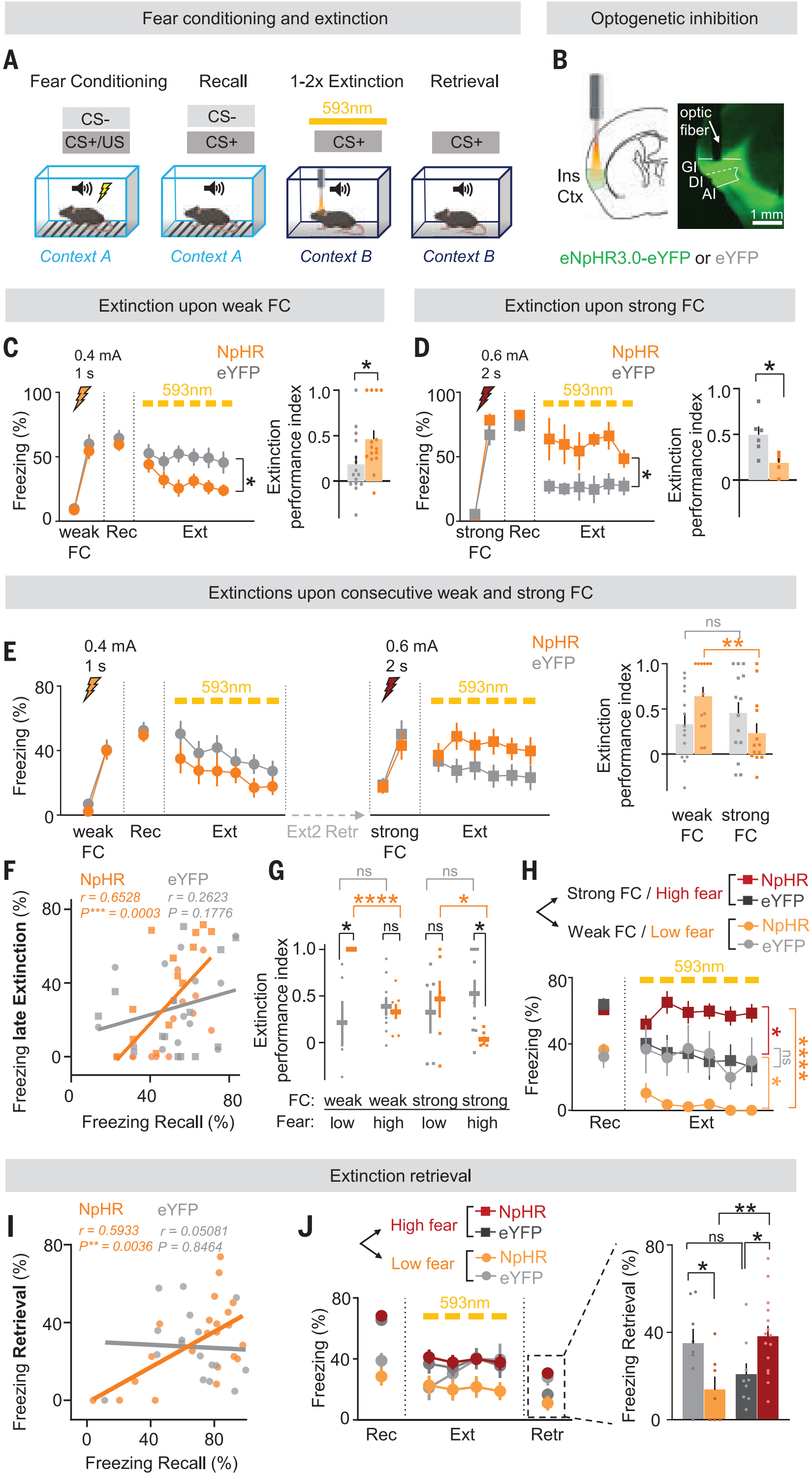
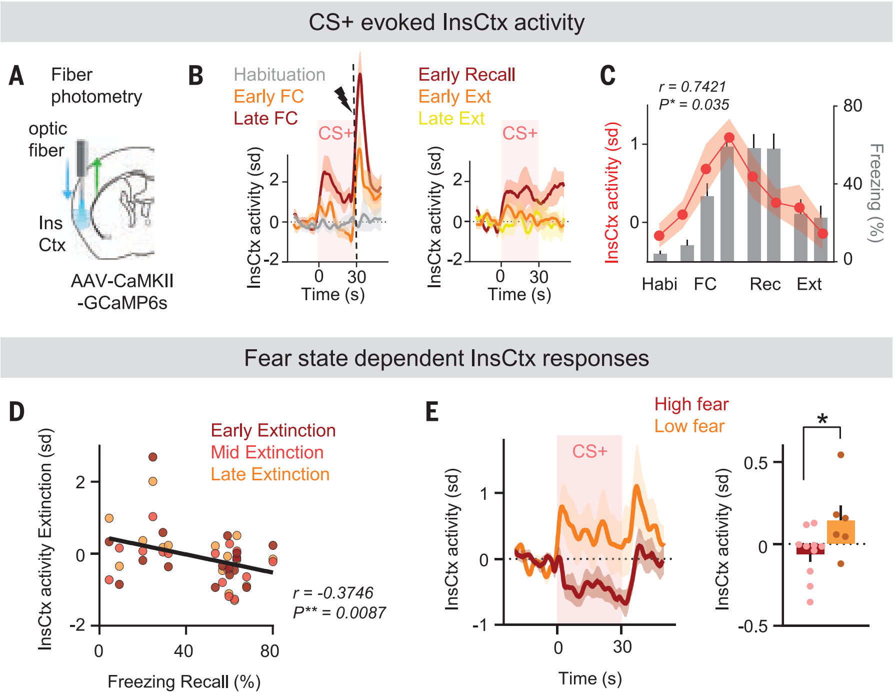


**Fig. 1. The InsCtx balances fear extinction and maintenance.** (**A**) The FC and extinction paradigm. (**B**) Strategy to inhibit the InsCtx. GI, granular InsCtx; DI, dysgranular InsCtx; AI, agranular InsCtx. (**C**) Extinction upon weak FC (left) and extinction performance indices (right) in *N* = 15 NpHR and *N* = 15 eYFP mice. Rec, recall; Ext, extinction. (**D**) Extinction upon strong FC (left) and extinction performance indices (right) in *N* = 5 NpHR and *N* = 6 eYFP mice. (**E** to **H**) Extinction upon consecutive weak and strong FC. Retr, retrieval; *r*, Pearson's correlation coefficient. Shown in (E) are the freezing behavior across all sessions (left) and extinction performance indices (right) in *N* = 13 NpHR and *N* = 14 eYFP mice. In (F), the correlation between fear level (freezing during recall) and freezing during late extinction upon weak (circles) or strong FC (squares) is shown. Shown in (G) are extinction performance indices of individual animals upon weak or strong FC depending on their fear level (high-fear mice: *N* = 7 NpHR, *N* = 9 eYFP; low-fear mice: *N* = 6 NpHR, *N* = 5 eYFP). Shown in (H) are the separation into the most affected groups (top; strong FC, high-fear mice: *N* = 7 NpHR, *N* = 9 eYFP; weak FC, low-fear mice: *N* = 6 NpHR, *N* = 5 eYFP) and freezing behavior of these groups (bottom). (**I** and **J**) The effect of InsCtx inhibition on extinction retrieval. In (I), the correlation between fear levels and freezing during extinction retrieval (*N* = 22 NpHR and *N* = 17 eYFP mice) is shown. Shown in (J) are the separation of the same mice into two groups according to their fear levels (top left; high fear: *N* = 15 NpHR, *N* = 9 eYFP; low fear: *N* = 7 NpHR, *N* = 8 eYFP), freezing behavior of high- and low-fear animals (bottom left), and quantification of freezing during retrieval (right). For (C) to (E), (H), and (J), filled symbols are average freezing values pooled from two CS+ presentations and four CS+ for recall. For behavior and protocol of the entire experiments, see figs. S3 and S4. The statistical tests used were as follows: for the left panels of (C) to (E), (H), and (J), two-way repeated measures analysis of variance (ANOVA); for the right panels of (C), (D), (G), and (J), two-tailed unpaired *t* tests; and for the right panel of (E), two-tailed paired *t* test. Detailed statistical results are shown in table S1. For all panels in which they appear, $*P \le 0.05$, $**P < 0.01$, $****P < 0.0001$, ns is not significant, and error bars represent mean ± SEM.

persisted during recall and eventually decreased back to baseline levels with progressing extinction learning (Fig. 2, B and C). InsCtx responsiveness to the CS+ and the amount of CS+ evoked freezing averaged across all animals were positively correlated throughout the different sessions of the paradigm (Fig. 2C). However, InsCtx responses did not increase to the same degree in response to a control tone, which was never paired with an aversive outcome (CS−; fig. S5, A and B).

Although these data revealed that InsCtx activity scales with the certainty to which the CS+ predicted the aversive outcome, they did not explain the observed paradoxical effects of InsCtx inhibition on extinction learning. We thus assessed InsCtx CS+ responsiveness in relation to individual fear levels. Contrary to the observation that CS+ responses in the InsCtx correlate to overall freezing across the fear and extinction paradigm (Fig. 2C), CS+ responses throughout extinction were negatively correlated to individual fear levels: Animals with higher recall fear exhibited lower CS+ evoked

**Fig. 2. InsCtx CS+ responses are modulated by fear state.** (**A**) Fiber photometry strategy. (**B**) CS+ evoked InsCtx activity during the fear conditioning and extinction paradigm (*N* = 16 mice). Lines represent average *z*-scored fluorescence traces ($\Delta F/F_0$) normalized to 30 s preceding each CS+. Habituation (Habi): first CS+; early-late FC: first-last CS+; early recall: first CS+; early-late extinction: first-last CS+. The dashed line in the left panel indicates the timing of the footshock (only delivered during FC). sd, standard deviation. (**C**) Average CS+ evoked InsCtx activity (red) and freezing (gray) during 30 s of CS+ presentations across the experiment [average of two CS+ presentations, except for mid-FC (one CS+)]. (**D**) Correlation between fear level and InsCtx activity during early, mid, and late extinction for each animal. (**E**) Mean CS+ evoked InsCtx activity during all CS+ presentations throughout extinction in high-fear (*N* = 10 mice) versus low-fear (*N* = 6 mice) animals (left); lines represent average *z*-scored fluorescence traces (ΔF/F0) normalized to 30 s preceding each CS+. The quantification of the same data using two-tailed unpaired *t* tests is shown on the right. Detailed statistical results are shown in table S1. For all panels in which they appear, $^*P \leq 0.05$, ns is not significant, and bars and shaded areas represent SEM.



InsCtx responses, and animals with lower recall fear responded with higher CS+ evoked InsCtx responses (Fig. 2D). Given the bidirectional effects of InsCtx inhibition on extinction learning, we again divided animals into high- and low-fear groups (fig. S5C). Consistent with the opposite behavioral changes upon InsCtx inhibition in high- versus low-fear animals, InsCtx activity increased upon CS+ presentations in low-fear animals, but exhibited an inverted response, namely an activity decrease, in high-fear animals (Fig. 2E).

Which process results in lowered InsCtx responses to the CS+ in high-fear compared with low-fear animals? High-fear animals exhibited substantially more freezing. Data from humans and other species suggest that freezing slows heart rate (*28*–*31*). The InsCtx receives rich interoceptive information from the body, including from the heart (*11*, *13*, *32*). We therefore tested how heart rate varies throughout fear conditioning and extinction. We used pulse oximetry in freely moving mice and analyzed heart rate deviations from baseline throughout all phases of the paradigm (Fig. 3, A to D; and fig. S6A). FC resulted in strong heart rate deviations that increased with successive CS+ and US pairings. Although painful footshocks resulted in heart rate increases, heart rates subsequently decelerated and were lowest during and after the CS+ presentations. During recall and extinction, strong fluctuations were triggered by the onset of the CS+ presentations but progressively diminished with ongoing extinction learning (Fig. 3, A to D). Overall, heart rate deviations were greater during FC and recall and decreased during extinction learning back to baseline levels (Fig. 3B). Notably, heart rates varied more strongly in high-fear than in low-fear animals (Fig. 3, A, C, and D).

To determine whether deviations in heart rate were reflected in InsCtx activity changes, we combined pulse oximetry and fiber photometry (Fig. 3, E to K; and fig. S6B). Heart rate deviation and InsCtx activity were not correlated during the initial phases of each session. We observed strong correlations upon the presentation of the first CS+ in FC, recall, and extinction (Fig. 3, E and F). Because the presentation of the CS+ marks the onset of a shift into a higher fear state, we hypothesized that the onset of the correlation between InsCtx activity and heart rate could be explained by changes in bodily feedback during fear expression. Indeed, heart rate decelerated with freezing onset, as reported previously (*28*–*31*) (Fig. 3G). Similarly, InsCtx activity decreased with the onset of freezing but was not correlated to speed of movement (fig. S6C), supporting the role of the InsCtx in interoception (Fig. 3H). Freezing-induced heart rate and InsCtx activity decreases were equally pronounced in high- and low-fear animals (Fig. 3, G and H) and thus could not explain the observed differences in InsCtx activities between these two groups (Fig. 2F). However, during extinction, high-fear animals exhibited overall more frequent and longer freezing bouts in response to CS+ presentations than low-fear animals (Fig. 3I). CS+ evoked InsCtx responses were negatively correlated with acute freezing during extinction (Fig. 3J). Analyzing high- and low-fear animals separately highlighted that InsCtx activity decreased during CS+ presentations in high-fear animals but increased in low-fear animals (Fig. 3K).

Our data suggest that InsCtx reactivity to a fear-associated cue results from the integration of the predictive value of the CS+ modulated by negative bodily feedback signals during acute freezing. Further, the different CS+ responsiveness in high- and low-fear states could be a product of stronger negative bodily feedback signals received in the InsCtx during frequent freezing, which is characteristic of high-fear animals. To test this hypothesis, we directly interfered with the communication between body and brain by means of vagus nerve stimulation (VNS). The vagus nerve is the main pathway carrying interoceptive information from the body to the brain (*33*–*35*) and is thought to relay bodily signals through a multisynaptic pathway to the visceral InsCtx (*11*, *13*, *36*) (Fig. 4A). We thus combined VNS (fig. S7B) with fiber photometry in the InsCtx (Fig. 4B). VNS increased the activity of the InsCtx when applied in a neutral context, confirming our approach (Fig. 4C). Analogous to our optogenetic inhibition experiments, we applied VNS during the CS+ presentations throughout extinction learning (Fig. 4D and fig. S7A). First, we tested whether we could override bodily feedback signals that modulate InsCtx responses during freezing. Indeed, during VNS application, InsCtx activity increased with the onset of a freezing episode,

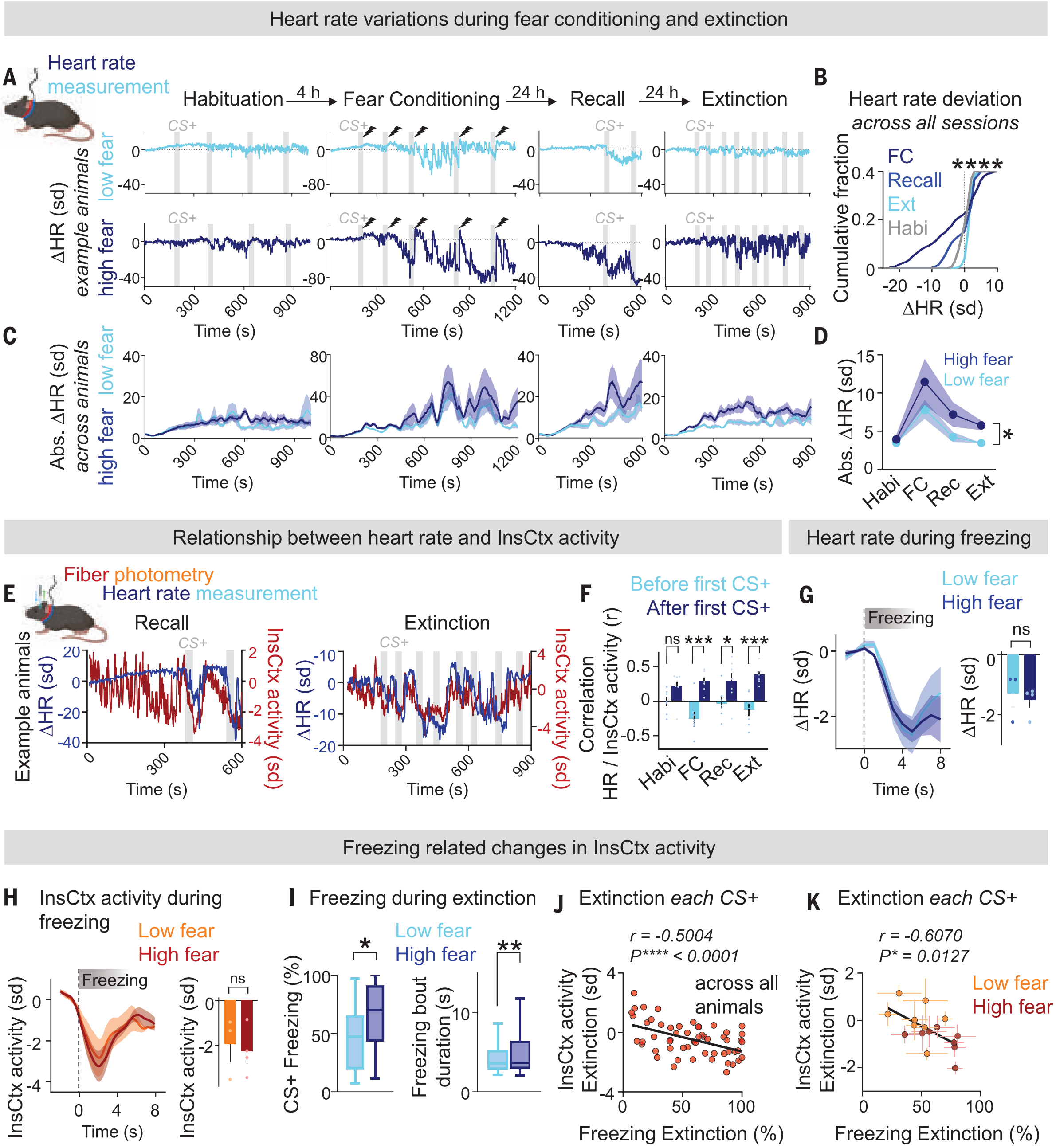


**Fig. 3. Fear expression modulates heart rate and InsCtx activity.** (**A** to **D**) Heart rate (HR) measurements using pulse oximetry during the FC and extinction paradigm. Shown in (A) are heart rate deviations (normalized to the first 100 s of each session) during each session of the paradigm in exemplary low-fear (top, light blue) and high-fear animals (bottom, dark blue). Shown in (B) are cumulative fractions of normalized heart rate deviations for each session of the paradigm, averaged across all animals (*N* = 15 mice). In (C), absolute heart rate deviations of low-fear (*N* = 7 mice) and high-fear (*N* = 8 mice) animals across all four sessions of the experiment are shown. Quantification of the data in (C) is shown in (D). (**E** to **K**) Combination of fiber photometry recordings in the InsCtx with heart rate measurements. In (E), heart rate deviation (dark blue) and InsCtx activity (dark red) during recall (left) and extinction (right) in individual exemplary animals are shown. In (F), correlations between heart rate deviation and InsCtx activity before and after the first CS+ presentation in each session (*N* = 7 mice) are shown. Shown in (G) are heart rate during freezing normalized to 2 s preceding freezing onset and averaged across high-fear (*N* = 4 mice) and low-fear (*N* = 3 mice) animals (left) and quantification of the same data (right). In (H), InsCtx activity during freezing normalized to 2 s preceding freezing onset and averaged across high- and low-fear animals (left) and quantification of the same data (right) are shown. Shown in (I) are the average freezing to each CS+ presentation during extinction [left; *n* = 24 CS+ for low-fear animals; *n* = 32 CS+ for high-fear animals (total of eight per animal)] and average freezing bout duration in low-fear (*n* = 136 freezing episodes) and high-fear (*n* = 132 freezing episodes) animals (right). In (J), the correlation between freezing and InsCtx activity for each CS+ presentation during extinction for all animals (*N* = 7 mice) is shown. In (K), the correlation between freezing and InsCtx activity to each CS+ presentation during extinction averaged across high-fear (*N* = 4 mice) and low-fear (*N* = 3 mice) animals is shown. The statistical tests used were as follows: for (B), Kolmogorov-Smirnoff tests; for (D), two-way repeated measures ANOVA; for (F), two-way repeated measures ANOVA with Bonferroni correction; and for (G) to (I), two-tailed unpaired *t* tests, with box-and-whisker plots in the style of Tukey. Detailed statistical results are shown in table S1. For all panels in which they appear, $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, $****P < 0.0001$, ns is not significant, and error bars and shaded areas represent SEM.

whereas, consistent with our previous findings, freezing resulted in InsCtx activity decreases in periods when no VNS was applied (Fig. 4E). Notably, VNS resulted in an inversion of the previously observed correlation between CS+ evoked InsCtx activity and freezing, such that in the presence of VNS, InsCtx activity was positively correlated to CS+-evoked freezing levels (Fig. 4F). Separating the VNS cohort into high- and low-fear animals (fig. S7C) revealed that inversely to the observations in nonmanipulated animals (Fig. 2F), upon VNS, high-fear animals responded with an InsCtx activity increase, whereas low-fear animals exhibited an activity decrease, in response to the CS+ presentations (Fig. 4G).

Finally, we addressed how VNS affects extinction learning. Consistent with the idea that bodily feedback transmitted by the vagus nerve provides the necessary bodily feedback signals to gate fear extinction learning, we found that VNS had fear state–dependent effects on extinction performance. Similar to InsCtx inhibition, under VNS, high-fear animals exhibited impaired extinction performance, whereas low-fear animals exhibited an extinction facilitation, both acutely during extinction and at retrieval of extinction memories (Fig. 4, H to J, and fig. S7D).

Our findings reveal that fear is actively maintained within balance by the InsCtx, which can either boost or weaken extinction learning depending on the fear state of the subject. Freezing, a shared means of fear expression across species (*28*, *37*), occurred together with heart rate decelerations, which are transmitted by the vagus nerve to the InsCtx and attenuate its responses to aversive cues. These findings highlight a neural mechanism by which freezing gains a coping function: It dampens aversive signaling in the InsCtx through bodily feedback. Our data suggest that disrupting this coping function may result in exaggerated fear memories and identify bodily changes that occur during freezing as an essential part of emotion regulation. Interestingly, stronger heartbeat perception in humans has been linked to pathological fear

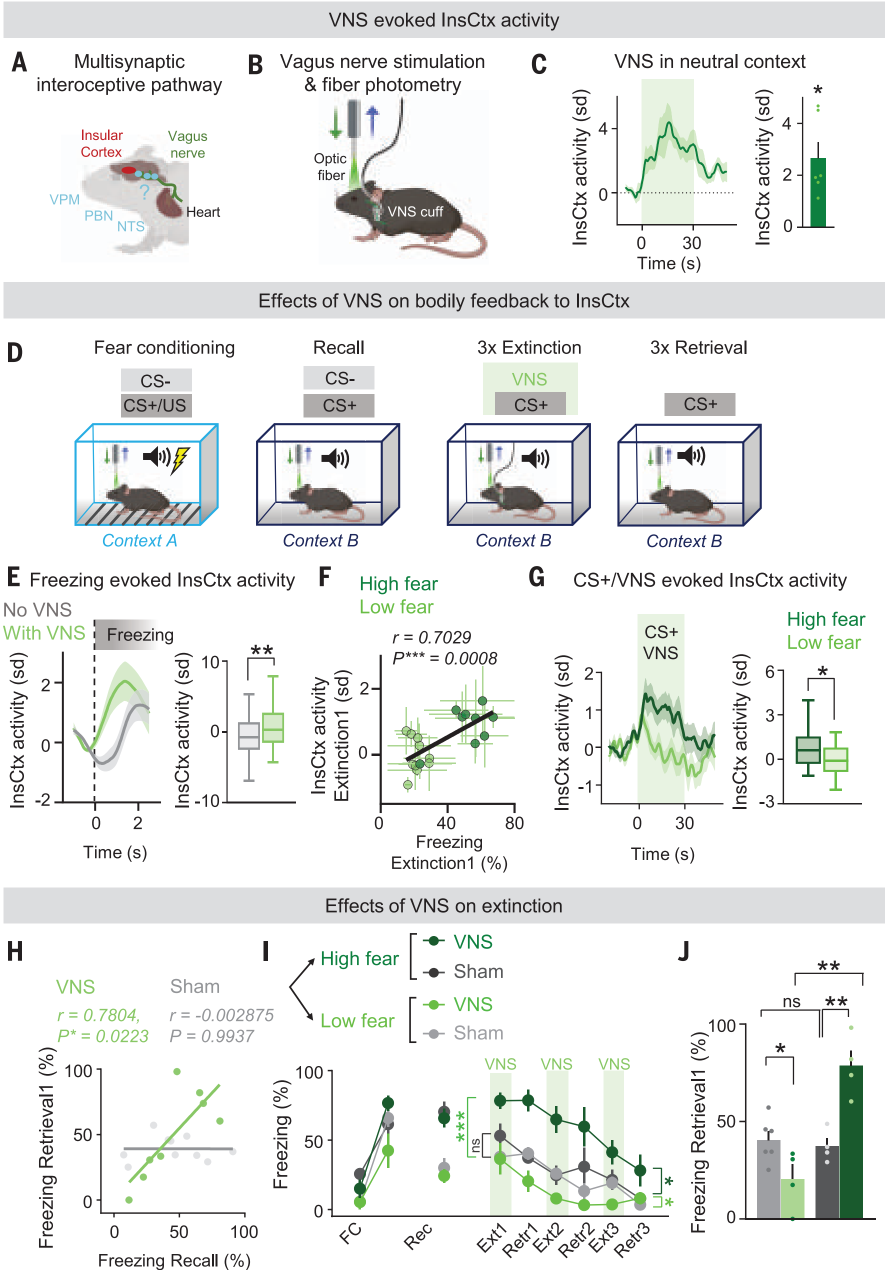


**Fig. 4. VNS disrupts fear balance.** (**A**) Proposed multisynaptic interoceptive pathway connecting cardiac activity to InsCtx activity. VPM, ventral posteromedial nucleus of the thalamus; PBN, parabrachial nucleus; NTS, nucleus of the solitary tract. (**B** to **G**) Fiber photometry in the InsCtx combined with VNS. Shown in (C) are VNS-evoked InsCtx activity (average trace of two stimulations) under neutral conditions (left) and quantification of the same data (right; *N* = 6 mice). In (D), VNS was applied during CS+ presentations throughout three consecutive extinction sessions. Shown in (E) are the average InsCtx activity during freezing events occurring within (*n* = 64 events) or outside (*n* = 100 events) VNS (left) and quantification of the same data (right). In (F), the correlation between freezing and InsCtx activity to each CS+ presentation during extinction (session 1) averaged across high-fear (*N* = 3 mice) and low-fear (*N* = 3 mice) animals is shown. Shown in (G) are InsCtx activity during all 10 CS+ presentations combined with VNS during extinction in the same high- and low-fear animals as in (F) (left) and quantification of the same data [right; *n* = 30 CS+ for high-fear mice; *n* = 30 CS+ presentations for low-fear mice (total of 10 CS+ per mouse)]. (**H** to **J**) Effects of VNS on fear extinction. In (H), the correlation between fear level and freezing during first retrieval of VNS (*N* = 8 mice) and sham control (*N* = 10 mice) animals is shown. Shown in (I) are the separation of the same mice as in (H) into groups according to their fear levels (top; *N* = 4 high-fear VNS, *N* = 4 high-fear sham, *N* = 4 low-fear VNS, *N* = 6 low-fear sham) and freezing behavior in both groups during FC and extinction (bottom). One data point represents the average of two CS+ for FC, recall, and retrievals and 10 CS+ for extinctions. For the protocol of the entire experiment, see fig. S7. Shown in (J) is the quantification of freezing during retrieval. The statistical tests used were as follows: for (C), one-sample *t* test with hypothetical mean of zero; for (E) right, (G) right, and (J), two-tailed unpaired *t* tests; and for (I), two-way repeated measures ANOVA. Detailed statistical results are shown in table S1. For all panels in which they appear, box-and-whisker plots are in the style of Tukey, $^{*}P \leq 0.05$, $^{**}P \leq 0.01$, ns is not significant, and error bars and shaded areas represent SEM.

and anxiety (*7*) and is known to engage the insular and opercular regions as key structures of the interoceptive brain (*9*, *32*).

Our data further suggest the integration of two opposite signals within the InsCtx: prediction of threat by fear-associated cues and negative feedback signals from the body. Future work is needed to address how interoception may modulate other parts of the fear circuitry, such as the amygdala or prefrontal cortex, which have also been reported to respond to bodily signals (*38*–*41*).

Although the role of the InsCtx as a regulator of bodily homeostasis has been suggested for a long time (*10*, *42*–*44*), we show in this study that the InsCtx also serves as homeostatic regulator of emotion. The InsCtx may detect deviations from established set points of adaptive functioning, for example, responding to fear as a deviation from an affective and bodily set point, which has already been suggested previously (*22*, *42*–*45*). Additionally, to reestablish affective homeostasis, the InsCtx may provide teaching signals to downstream effector systems, such as the known extinction circuitry (*46*–*48*) to which it is heavily connected (*11*, *12*). Future work should address how other types of bodily feedback to the InsCtx—such as hunger, craving, and gut- or breathing-related signals (*13*, *44*, *45*, *49*, *50*)—may influence InsCtx function in regulating affective states.

Although our results indicate that the InsCtx may be dispensable for extinction in subjects with moderate fear levels, its gating function becomes important in high- or low-fear subjects that are experiencing extreme situations. These findings may relate to the dual role of the InsCtx in anxiety and addiction disorders (*22*, *49*, *51*–*53*), which are characterized by opposing ways to deal with threat and harm (*49*). Furthermore, people suffering from anxiety disorders exhibit hyperactivity in the InsCtx when compared with healthy controls (*22*), possibly owing to the lack of negative feedback from the body.

Although VNS facilitates fear extinction in rats (*54*) and is tested in clinical studies to treat anxiety disorders in humans (*55*), our results reveal that VNS may worsen extinction outcomes under certain conditions. This is in agreement with a recent study (*54*) and emphasizes the need to individually adjust VNS therapy. Despite this caveat, our study highlights the potency of engaging a bodily route in emotion regulation. Given the overall poor accessibility of the human brain for targeted therapeutic interventions, the detailed description of this bodily feedback mechanism may bear enormous therapeutic potential for human anxiety disorders.

## REFERENCES AND NOTES

1. D. Mobbs, J. J. Kim, *Curr. Opin. Behav. Sci.* **5**, 8–15 (2015).
2. O. Stiedl, J. Spiess, *Behav. Neurosci.* **111**, 703–711 (1997).
3. I. Van Diest, M. M. Bradley, P. Guerra, O. Van den Bergh, P. J. Lang, *Biol. Psychol.* **80**, 212–217 (2009).
4. S. Bagur *et al.*, *Nat. Commun.* **12**, 2605 (2021).
5. M. Maschke *et al.*, *J. Neurol. Neurosurg. Psychiatry* **72**, 116–118 (2002).
6. M. P. Paulus, M. B. Stein, *Biol. Psychiatry* **60**, 383–387 (2006).
7. A. J. Willem Van der Does, M. M. Antony, A. Ehlers, A. J. Barsky, *Behav. Res. Ther.* **38**, 47–62 (2000).
8. L. Quadt, H. D. Critchley, S. N. Garfinkel, *Ann. N. Y. Acad. Sci.* **1428**, 112–128 (2018).
9. S. N. Garfinkel, H. D. Critchley, *Trends Cogn. Sci.* **20**, 34–46 (2016).
10. A. D. Craig, *Nat. Rev. Neurosci.* **3**, 655–666 (2002).
11. G. V. Allen, C. B. Saper, K. M. Hurley, D. F. Cechetto, *J. Comp. Neurol.* **311**, 1–16 (1991).
12. D. A. Gehrlach *et al.*, *eLife* **9**, e55585 (2020).
13. D. F. Cechetto, C. B. Saper, *J. Comp. Neurol.* **262**, 27–45 (1987).
14. J. P. Casanova *et al.*, *Behav. Brain Res.* **296**, 70–77 (2016).
15. C. Sehlmeyer *et al.*, *PLOS ONE* **4**, e5865 (2009).
16. A. R. Foilb, J. G. Flyer-Adams, S. F. Maier, J. P. Christianson, *Neurobiol. Learn. Mem.* **134**, 317–327 (2016).
17. J. P. Casanova, M. Aguilar-Rivera, M. L. Á. Rodríguez, T. P. Coleman, F. Torrealba, *J. Neurophysiol.* **120**, 1906–1913 (2018).
18. N. M. Brydges *et al.*, *PLOS ONE* **8**, e54197 (2013).
19. M. A. Fullana *et al.*, *Mol. Psychiatry* **21**, 500–508 (2016).
20. J. A. Greco, I. Liberzon, *Neuropsychopharmacology* **41**, 320–334 (2016).
21. D. Kargl *et al.*, *eLife* **9**, e60336 (2020).
22. A. Etkin, T. D. Wager, *Am. J. Psychiatry* **164**, 1476–1488 (2007).
23. S. S. Khalsa *et al.*, *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 501–513 (2018).
24. T. Klucken, O. Kruse, J. Schweckendiek, R. Stark, *Front. Behav. Neurosci.* **9**, 132 (2015).
25. D. W. Grupe, J. B. Nitschke, *Nat. Rev. Neurosci.* **14**, 488–501 (2013).
26. K. Preuschoff, S. R. Quartz, P. Bossaerts, *J. Neurosci.* **28**, 2745–2752 (2008).
27. L. A. A. Gunaydin *et al.*, *Cell* **157**, 1535–1551 (2014).
28. K. Roelofs, *Philos. Trans. R. Soc. London Ser. B* **372**, 20160206 (2017).
29. M. A. Hagenaars, M. Oitzl, K. Roelofs, *Neurosci. Biobehav. Rev.* **47**, 165–176 (2014).
30. L. C. Schenberg, E. C. Vasquez, M. B. da Costa, *Brain Res.* **621**, 50–58 (1993).
31. J. Iwata, J. E. LeDoux, *Behav. Neurosci.* **102**, 66–76 (1988).
32. H. D. Critchley, S. Wiens, P. Rotshtein, A. Öhman, R. J. Dolan, *Nat. Neurosci.* **7**, 189–195 (2004).
33. H. R. Berthoud, W. L. Neuhuber, *Auton. Neurosci.* **85**, 1–17 (2000).
34. R. Câmara, C. J. Griessenauer, in *Nerves and Nerve Injuries*, R. S. Tubbs, E. Rizk, M. M. Shoja, M. Loukas, N. Barbaro, R. J. Spinner, Eds. (Academic Press, 2015), pp. 385–397.
35. S. Ito, *Neurosci. Lett.* **148**, 151–154 (1992).
36. M. Kalia, J. M. Sullivan, *J. Comp. Neurol.* **211**, 248–265 (1982).
37. A. D. Zych, N. Gogolla, *Curr. Opin. Neurobiol.* **68**, 57–66 (2021).
38. R. C. Frysinger, R. M. Harper, *Electroencephalogr. Clin. Neurophysiol.* **72**, 463–470 (1989).
39. J. Liu, L. Lin, D. V. Wang, *J. Neurosci.* **41**, 1080–1091 (2021).
40. A. Adhikari *et al.*, *Nature* **527**, 179–185 (2015).
41. L. B. M. Resstel, F. M. A. Corrêa, *Auton. Neurosci.* **126-127**, 130–138 (2006).
42. G. Egan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15241–15246 (2003).
43. Y. Livneh *et al.*, *Nature* **546**, 611–616 (2017).
44. Y. Livneh *et al.*, *Neuron* **105**, 1094–1111.e10 (2020).
45. D. A. Gehrlach *et al.*, *Nat. Neurosci.* **22**, 1424–1437 (2019).
46. P. Tovote, J. P. Fadok, A. Lüthi, *Nat. Rev. Neurosci.* **16**, 317–331 (2015).
47. C. Herry *et al.*, *Eur. J. Neurosci.* **31**, 599–612 (2010).
48. G. G. Calhoon, K. M. Tye, *Nat. Neurosci.* **18**, 1394–1404 (2015).
49. N. H. Naqvi, N. Gaznick, D. Tranel, A. Bechara, *Ann. N. Y. Acad. Sci.* **1316**, 53–70 (2014).
50. M. Aguilar-Rivera, S. Kim, T. P. Coleman, P. E. Maldonado, F. Torrealba, *Sci. Rep.* **10**, 21642 (2020).
51. J. Downar, D. M. Blumberger, Z. J. Daskalakis, *Trends Cogn. Sci.* **20**, 107–120 (2016).
52. M. Goodkind *et al.*, *JAMA Psychiatry* **72**, 305–315 (2015).
53. H. Namkung, S.-H. Kim, A. Sawa, *Trends Neurosci.* **40**, 200–207 (2017).
54. R. R. Souza *et al.*, *Exp. Neurol.* **341**, 113718 (2021).
55. M. S. George *et al.*, *Brain Stimul.* **1**, 112–121 (2008).
56. A. S. Klein, N. Dolensek, C. Weiand, N. Gogolla, Source data for "Fear balance is maintained by bodily feedback to insular cortex in mice." Zenodo (2021);

## ACKNOWLEDGMENTS

We thank members of the Gogolla laboratory, C. Herry, J. J. Letzkus, and R. Portugues for technical guidance, discussions, and feedback; F. Lyonnaz and E. Cho for technical assistance; and M. Malezieux for feedback on the manuscript. **Funding:** This work was supported by the Max-Planck Society, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-2017-STG, grant agreement n° 758448 to N.G.), and the ANR-DFG project "SAFENET" (ANR-17-CE37-0021). **Author contributions:** N.G. and A.S.K. conceived the project and designed the experiments. A.S.K. performed all experiments and analyses. N.D. provided technical assistance and helped with analyses and programming. C.W. helped with experiments, controls, and histology. N.G. wrote the manuscript with the assistance of all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript and the supplementary materials and at Zenodo (*56*). Viruses were packaged at the University of North Carolina (UNC) Vector Core or at Addgene and are made available under material transfer agreements. Illustrations containing drawings were created with BioRender.com.

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj8817
Materials and Methods
Supplementary Text
Figs. S1 to S7
Table S1
References (*57*–*60*)

8 June 2021; accepted 21 September 2021
10.1126/science.abj8817

## FIVE-YEAR DAVID COHEN (DC) BIOMAGNETISM FELLOWSHIP

The Athinoula A. Martinos Center for Biomedical Imaging invites candidates to submit applications for a five-year **David Cohen (DC) Biomagnetism Fellowship**. The David Cohen Postdoctoral Fellow will develop experimental and analytical tools to measure ultra low-frequency or dc magnetic fields of biological origin and to apply these methods and tools to basic and clinical research studies of neurological, psychiatric, or other disorders.

The Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH) is one of the world's premier research centers devoted to the development and application of advanced biomedical imaging technologies. The Center is part of the Department of Radiology at Massachusetts General Hospital and is affiliated with both Harvard Medical School (HMS) and Massachusetts Institute of Technology (MIT). Our mission is to revolutionize imaging for the advancement of science and medicine. Inspire, educate, and inform. The Martinos Center provides an exceptionally rich multidisciplinary environment for collaborations and mentoring at all levels of training. More information can be found at the Martinos Center website **https://www.martinos.org**.

The endowed DC Fellowship is intended to provide salary support to a senior postdoctoral fellow for five years as they develop an independent research program and transition to junior faculty. To be eligible, a candidate must hold a Ph.D. and must have prior research experience in bioelectric or biomagnetic fields.

The fellowship will meet the NIH salary levels for Postdoctoral Trainees and Fellows, see **https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-049.html**.

To apply, applicants should send a single PDF to **dc_committee@nmr.mgh.harvard.edu** containing:

1. A cover letter with contact information
2. A CV in an NIH biosketch format
3. A four-page proposal including one page of specific aims, one page of background and significance, and two pages describing the research plan and experimental approach. The total length with citations can exceed the four-page limit.
4. Applicants should also arrange to have three letters of recommendation sent directly to **dc_committee@nmr.mgh.harvard.edu**. The letters should be from individuals familiar with the applicant's research and potential to transition as an independent investigator.

Instructions, a sample, and a blank format page for an NIH Biosketch can be found at (**https://grants.nih.gov/grants/forms/biosketch.htm**). Please follow the Postdoctoral Fellowship biosketch format under Biographical Sketch Format Page (non-fellowship) Due Dates on/after **January 25, 2022**. An eRA Commons User Name is not required.

Should you have any questions regarding the fellowship, please do not hesitate to contact us by email: **dc_committee@nmr.mgh.harvard.edu**

The deadline for applications and letters of recommendation is **31 January, 2022**.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS
Confused about your next career move?
Download Free Career Advice Booklets!
ScienceCareers.org/booklets
CAREER TRENDS

EPFL

# FACULTY POSITIONS IN NEUROSCIENCE

## AT THE ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL)

The Ecole polytechnique fédérale de Lausanne (EPFL) invites applications for two faculty positions at the Tenure Track **Assistant Professor** level in the **Brain Mind Institute of the School of Life Sciences**. We seek candidates committed to research in the broad fields of neuroscience, neuroengineering and neurotechnology with interest in brain (dys)function, neurodegeneration, regeneration, stem cells and organoids studied from molecular, genetic, cellular, neural circuit, systems, imaging, computational and/or behavioral perspectives.

The candidates should also demonstrate a strong commitment to excellence in teaching at the undergraduate and graduate levels.

The new professors will join a dynamic EPFL community that is driven by scientific discovery and technological advances and operate in state-of-the-art labs across EPFL facilities in Lausanne, Geneva, Fribourg and Sion. We anticipate the new faculty members will take advantage of scientific interactions with the School of Life Sciences, School of Engineering, School of Computer and Communication Sciences, Center for Neuroprosthetics, Blue Brain Project, Center for Imaging, Center for Intelligent Systems, Fribourg Translational Neuroscience platform, the Universities of Lausanne and Geneva and their associated hospitals, and the Swiss Rehabilitation Clinic in Sion. The School of Life Sciences currently has five broad strategic aims: Neuro-X [science, technology, computation], organoid bioengineering, biological data science, structural biology, and environmental biology.

Commitment to open science and data dissemination will be positively valued.

Significant start-up resources, research budget, and state-of-theart research infrastructure are available. Salaries and benefits are internationally competitive.

Applications should include a cover letter, a curriculum vitae, a list of publications, concise statements of research and teaching interests as well as the contact information of at least three references who are ready to supply their letter upon request.

Applications must be submitted to: **https://facultyrecruiting.epfl.ch/position/35848196**

Formal evaluation of candidates will begin on **15th December, 2021**.

For enquiries regarding the positions, please contact:
Prof. Carl Petersen, Chair of the Search Committee
E-mail: **neurosearch2022@epfl.ch**

Additional information may be found by consulting the web sites: **www.epfl.ch**, **www.epfl.ch/schools/sv** and **www.epfl.ch/schools/sv/bmi**

*EPFL is an equal opportunity employer and a family friendly university. It is committed to increasing the diversity of its faculty, and strongly encourages women to apply.*

---

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

# Tenure-Track Assistant Professor Neurogenetics

**Position Description Summary:**
The Department of Biological Sciences at the George Washington University (GWU) invites applicants for a tenure-track faculty position in Neurogenetics at the rank of **Assistant Professor**, to begin in Fall 2022. Our department's research interests focus primarily on non-mammalian systems, and the successful candidate will be expected to establish and maintain an externally funded research program at the intersection of experimental neurobiology and genetics. The teaching expectations include a genetics course with a laboratory component. We particularly welcome candidates from groups that are historically underrepresented in biology and/or have demonstrated leadership toward building an equitable and inclusive scholarly environment.

To be considered, please complete an online faculty application (**https://www.gwu.jobs/postings/87426**). Review of applications will begin on **December 8, 2021** and continue until the position is filled, pending final budgetary approval. Employment offers are contingent on the satisfactory outcome of a standard background screening.

*The university is an Equal Employment Opportunity/Affirmative Action employer that does not unlawfully discriminate in any of its programs or activities on the basis of race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity or expression, or on any other basis prohibited by applicable law.*

---

Zertifikat seit 2007
audit familiengerechte hochschule

Technische
Universität
Braunschweig

The **Faculty of Electrical Engineering, Information Technology, Physics**, Institute of Applied Physics offers a

# University professorship (m/f/d) for "Robust Quantum Systems" (Salary-Gr. W3)

The person to be appointed competently represents the subject area of experimental physics "Robust Quantum Systems". The research focus is in the area of solid-state-based quantum physics or quantum technology. Possible thematic orientations can be in the area of hybrid structures as well as effectively two-dimensional materials with extraordinary electronic properties, which can serve as robust quantum systems with coupling to the environment for manipulation, detection and transfer of quantum states. We envision a thematic and methodological expansion of the research portfolio already established in Braunschweig in experimental and theoretical solid-state physics. In teaching, the professorship has to represent experimental physics in its full breadth.

In addition to scientific excellence, an important criterion will be the fit with the activities in the Cluster of Excellence QuantumFrontiers (QF) and the collaboration in the research centre Laboratory for Emerging Nanometrology (LENA). In QF and LENA the Physikalisch-Technische Bundesanstalt is incorporated as a partner. Another possibility is the active participation in the Quantum Computing Initiative of the State of Lower Saxony "Quantum Valley Lower Saxony" (QVLS).

Vacancy offer: **https://www.tu-braunschweig.de/en/vacancies**

Please send your application until **31$^{st}$ December 2021** using the **online form** of TU Braunschweig.

WORKING LIFE

*By* **Rassim Khelifa**

# The hard way

I signed in to our virtual lab meeting feeling especially nervous. I had volunteered to give a presentation about my background as a first-generation college student from Africa. For the first time in 8 years abroad as an international researcher, I planned to point out privileges that scientists from the Global South had to work hard to earn, but that my colleagues in Canada might take for granted. I did not know how my postdoc supervisor and fellow lab members would react. But I had decided it was worth taking a risk to raise awareness and perhaps spur action.

I grew up in Algeria and have had to pay an extra price throughout my scientific journey as a result. Unlike most wildlife researchers, I was disconnected from nature growing up. A civil war had made wild areas unsafe to visit. I encountered the natural world during biology field trips as an undergraduate student and quickly fell in love with it. I decided I wanted to become a wildlife biologist, a field that was in its infancy in my country.

I knew studying abroad would be key for my success, so I was thrilled when I won a scholarship to attend a U.K. university for graduate school. But there was a catch: I needed to take an English test. I was fluent in Arabic and French, the languages spoken in Algeria, but I didn't have a solid grasp of English. I studied hard but didn't score high enough to secure entrance. The university asked me to retake it, but I couldn't afford the cost of another test.

"It's important to help privileged scientists recognize their own privilege."

I stayed in Algeria to complete a master's degree instead. I kept learning English on my own and started to write scientific manuscripts about my work. Before submitting each paper, I sent it to researchers in Europe whom I'd never met, asking for feedback. It was the only way I could think of to receive language help. To my surprise, some of them sent responses and kindly revised my English.

Two years later, I earned a Ph.D. scholarship to study in Switzerland. I was excited about the opportunity. But it was discouraging to learn that the coursework I'd taken in Algeria was deemed insufficient. I had to spend a semester taking more than double the usual amount of credits.

Midway through my Ph.D., a professor remarked, "You are quite productive, but why don't you publish in the regular journals in our field?" Shocked, I responded, "What do you mean by regular journals?" He gave a few examples, all journals dominated by researchers from the Global North. I had been publishing my papers in specialized, low-impact journals. That counted as a great achievement in my home country, where even professors struggle to get published. Now, I realized that abroad, the scientific accomplishments I was proud of were perceived as below average at best.

The professor's comment spurred me to adjust my research to address issues of broader interest. By the time I graduated, I had started publishing in so-called "regular" journals, which helped me land a postdoc in Canada.

I told all this to members of my lab, hoping to give them some sense of the journey I'd taken to get there. I wanted them to know they were privileged to grow up speaking English and to have access to tremendous expertise and funding. To my relief, my supervisor and lab mates thanked me afterward. Their feedback helped me see my journey as more than a series of hurdles. It's also an intellectual asset, giving me a platform to speak up about global inequities.

Since my lab presentation, I've continued to share my story. I believe it's important to help privileged scientists recognize their own privilege—because that may lead to a shift in how they interact with others. For instance, wildlife biologists who travel to Africa to conduct research could seek collaborations with local researchers. And scientists with strong English skills could help edit the work of researchers from non–English speaking countries. It's not easy to find those kinds of volunteer opportunities, but I'm working to change that. I've started discussions with a preprint server about integrating an English language peer proofing system into its platform, which would connect authors who need language help with scientists who are in a position to provide it.

We can take steps to make the global scientific community more inclusive. And I want to be part of that push for change. ■

---

Rassim Khelifa is a postdoc at the University of British Columbia.

ILLUSTRATION: ROBERT NEUBECKER

Baylor College of Medicine®
GIVING LIFE TO POSSIBLE

### DEPARTMENT OF MOLECULAR AND HUMAN GENETICS AND DAN L. DUNCAN COMPREHENSIVE CANCER CENTER

## TENURED/TENURE TRACK FACULTY POSITIONS IN GENETICS/GENOMICS

The Department of Molecular and Human Genetics and the Dan L. Duncan Comprehensive Cancer Center (DLDCCC) at Baylor College of Medicine are seeking the following positions:

- **Genome Instability** – an individual for faculty appointment at rank appropriate for achievement and experience, working in any organism on problems in genomic instability, development of novel genetic/genomic tools or other fundamental genetic topics.

Successful candidates will have strong basic research programs related to genetic/genomic stability or instability, genome organization, genomics, genome engineering including, but not limited to DNA replication, repair, mutation, genome rearrangements, DNA damage response, mechanisms of heritability and evolution, studied in any organism from bacteria to human. This individual will join the Mechanisms in Cancer Evolution Program in the DLDCCC. However, outstanding individuals in any research area will be considered. Generous start-up support is available, and candidate will be put forward for a CPRIT award. The weekly Genome Instability Group meeting, supported by the Department and DLDCCC, includes labs from throughout the Texas Medical Center discussing work in progress.

- **Cancer Geneticist** – an individual for faculty appointment at rank appropriate for achievement and experience in cancer research. Applicants' research programs may focus on broad ranging topics in cancer genetic research including cancer genomics, mechanisms of cancer therapeutics, cancer model organisms, genome instability, epigenetics and gene expression, and others. Applicants with expertise in computational biology, in combination with basic and/or translational cancer research, are strongly encouraged to apply. Candidates will join a team of multidisciplinary research investigators studying cancer stem cell biology, genomics, epigenetics, and metabolic aberrations in cancer.

Both appointments will be at the **Assistant**, **Associate**, or **Full Professor** level depending on experience.

The Department's research interests include genomics, mechanisms of DNA replication and repair, mechanisms of transcription, mutation, DNA recombination, genomic instability and cancer, behavioral genetics, bioinformatics, gene therapy, gene structure and expression, bacterial evolution and antibiotics resistance and the biology of aging. Department research includes strengths in bacterial, yeast, Dictyostelium, worm, fly, mouse and human genetics.

Among genetics departments at U.S. medical schools, the Department of Molecular and Human Genetics at Baylor College of Medicine (**https://www.bcm.edu/departments/molecular-and-human-genetics**) ranks first in both number of grants and total funding from the National Institutes of Health (NIH). The Department includes basic and clinical research, a new joint venture diagnostic laboratory, long-standing association with an NIH human genome sequencing center, an NIH-funded graduate program, and residency/fellowship training in medical genetics. The Department has 70 primary tenured and tenure-track research faculty members and a total of 140 primary faculty members, who are engaged in a variety of missions including basic and translational research, clinical diagnostic services, and prenatal, pediatric, and adult clinical care.

The Dan L Duncan Comprehensive Cancer Center (DLDCCC) includes over 270 research members in seven different programs who bring in $180M in total cancer relevant funding including nearly $90M from NIH. Twelve state-of-the-art Shared Resources are supported by the DLDCCC and, in total, BCM has more than 30 core facilities supporting research (**https://www.bcm.edu/centers/cancer-center/research/shared-resources**).

Houston is the fourth largest US city, the most internationally diverse US city, "is widely considered to be one of the most culturally diverse cities in the world"—NY Times, and is on the New York Times' 50 Places To-Go list in the world for food and culture.

Curriculum vitae, a brief summary of research plans, along with the names, addresses, and phone numbers of at least three references to the following email address:
**mhgfacultyrecruits@bcm.edu.**

Department of Molecular and Human Genetics
Baylor College of Medicine
One Baylor Plaza, ABBR Room R830
Houston, TX 77030
Phone: 713-798-5443 Fax: 713-798-8515

*Equal Opportunity, Affirmative Action and Equal Access Employer*

---

Baylor College of Medicine®
GIVING LIFE TO POSSIBLE

### DEPARTMENT OF MOLECULAR AND HUMAN GENETICS

## FACULTY POSITIONS

Among genetics departments at U.S. medical schools, the Department of Molecular and Human Genetics at Baylor College of Medicine (**https://www.bcm.edu/departments/molecular-and-human-genetics/**) ranks first in both number of grants and total funding from the National Institutes of Health (NIH). The Department of Molecular and Human Genetics provides a fully integrated environment for physicians and basic scientists, promoting a cross-species approach to functional genetics and a commitment to technology transfer. Activities within the Department include clinical genetics, basic and clinical research, a joint venture diagnostic laboratory, long-standing association with an NIH large scale human genome sequencing centers, medical student teaching, an MS Genetic Counseling program, a Ph.D. graduate program, and residency/fellowship training in medical genetics.

The Department has over $100 million in total research funding, 70 primary tenured and tenure-track research faculty members and a total of 140 primary faculty members, who are engaged in a variety of missions including basic and translational research, clinical diagnostic services, and prenatal, pediatric, and adult clinical care. To expand our translational impact, we established a uniquely structured and governed joint venture diagnostic laboratory, Baylor Genetics, and a Baylor College of Medicine/Chinese University of Hong Kong Center for Medical Genetics in Hong Kong that together has expanded our clinical, diagnostic, and educational mission to a world-wide audience.

Faculty will have an opportunity to interact with the vigorous ABMGG residency and clinical laboratory fellowship training programs that include over 40 clinical and diagnostic laboratory faculty and over 40 genetic counseling faculty and staff. The training programs attract four to six individuals per year for resident training in categorical Medical Genetics, Combined Pediatrics/Medical Genetics, Combined Internal Medicine/Medical Genetics, and Maternal Fetal Medicine/Medical Genetics, as well as three to four individuals for ABMGG-approved clinical laboratory fellowship training in the following areas: clinical biochemical genetics and laboratory genetics and genomics.

The Department staffs clinical genetics programs at Texas Children's Hospital, Baylor St. Luke's Medical Center, Harris Health System, and the Michael E. DeBakey Veteran Affairs Medical Center.

The Department is seeking individuals for faculty appointments at rank appropriate for achievement and experience. We are currently recruiting for the following positions:

- **Physician Scientist Geneticist** – The Department is seeking MD or MD/PhD trained individuals who have clearly demonstrated their ability to conduct high-impact basic, translational and/or clinical research and compete for scientific funding in the area of medical genetics and model organism genetics. Start-up package commensurate for experience will be provided.
- **Clinical Geneticist** – The Department is seeking MD or MD/PhD trained, ABMGG eligible and/or certified physicians in the area of clinical genetics and medical biochemical geneticist.
- **Research Geneticis**t – The Department is seeking PhD trained basic and translational research geneticists (studying any model system), statistical geneticists, and computational geneticists who are focused on elucidating the underlying genetic and genomic mechanisms of rare and common diseases.

Appointments will be at the **Assistant**, **Associate**, or **Full Professor** levels depending on experience.

Qualified applicants should email a pdf version of their curriculum vitae and cover letter stating the position for which they are applying to the following email address:
**mhgfacultyrecruits@bcm.edu**

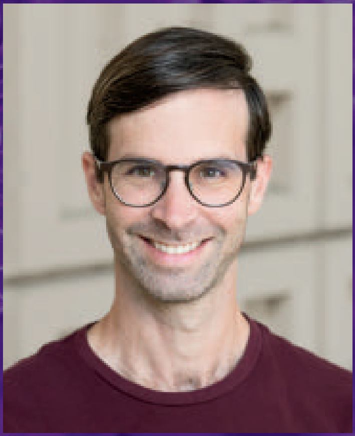
Department of Molecular and Human Genetics
Baylor College of Medicine
One Baylor College of Medicine, ABBR Room R830
Houston, TX 77030
Phone: 713-798-5443 Fax: 713-798-8515

*Equal Opportunity, Affirmative Action and Equal Access Employer*

# TRILLIONS OF MICROBES

# ONE ESSAY

The NOSTER *Science* Microbiome Prize is an international prize that rewards innovative research by investigators, under the age of 35, who are working on the functional attributes of the microbiota. The research can include any organism that has potential to contribute to our understanding of human or veterinary health and disease, or to guide therapeutic interventions. The winner and finalists will be chosen by a committee of independent scientist, chaired by a senior editor of *Science*. The top prize includes a complimentary membership to AAAS, an online subscription to *Science*, and $25,000 (USD). Submit you research essay today.



Eran Blacher, Ph.D.
2021 Winner

NOSTER | Science
MICROBIOME
PRIZE

**Apply by 1/24/22 at www.science.org/noster**

**Sponsored by Noster, Inc**

*The Frontiers of Medical Research*

A sponsored supplement to *Science*

Produced by the *Science*/AAAS Custom Publishing Office

Sponsored by

Icahn
School of
Medicine at
**Mount**
**Sinai**

A Sponsored Supplement to *Science*

Personalized Medicine

Exposomics

# The Frontiers of Medical Research

Artificial Intelligence &
Learning Health Systems

Health Equity

Cardiovascular Medicine

Microbiology

Biomedical Education

Surgery

Women's Health

Geriatrics & Palliative Care

Cancer

Immunology

Stem Cell Biology &
Regenerative Medicine

Genomic Medicine

Neuroscience

Maternal & Child Health

**Sponsored by**

Icahn School
of Medicine at
**Mount
Sinai**

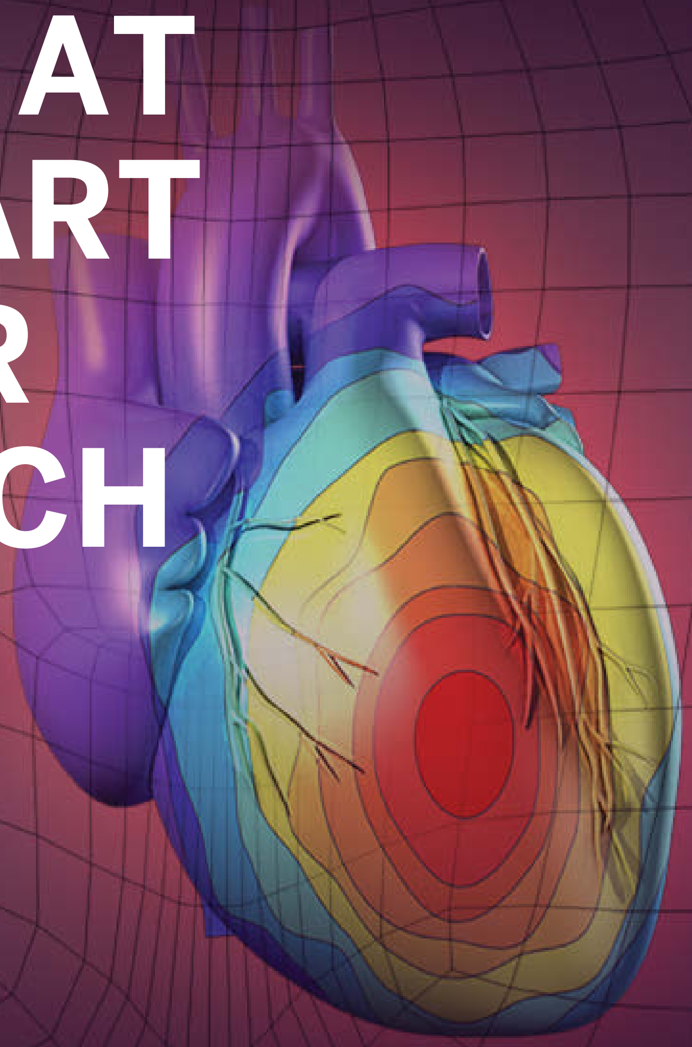
**Produced by the *Science*/AAAS
Custom Publishing Office**

Science
AAAS

ScienceTranslationalMedicine.org

PUT HUMAN HEALTH AT THE HEART OF YOUR RESEARCH

Submit your research:
**cts.ScienceMag.org**



Science
Translational
Medicine
AAAS

Twitter: @ScienceTM
Facebook: @ScienceTranslationalMedicine

# Table of Contents

A Sponsored Supplement to Science
Personalized Medicine
Exposomics
The Frontiers of Medical Research
Artificial Intelligence & Learning Health Systems
Health Equity
Cardiovascular Medicine
Microbiology
Biomedical Education
Surgery
Women's Health
Geriatrics & Palliative Care
Cancer
Immunology
Stem Cell Biology & Regenerative Medicine
Genomic Medicine
Neuroscience
Maternal & Child Health
Icahn School of Medicine at Mount Sinai
Science AAAS

This supplement was produced by the *Science*/AAAS Custom Publishing Office and sponsored by Icahn School of Medicine at Mount Sinai.

Materials that appear in this supplement have not been peer-reviewed nor have they been assessed by *Science*. Articles can be cited using the following format: [AUTHOR NAME(S)] [CHAPTER TITLE] in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), p. [xx-xx].

Editor: Sean Sanders, Ph.D.
Proofreader/Copyeditor: Bob French
Designer: JD Huntsinger

**Chris Hoag**
Sales Manager
choag@aaas.org
+1 (202) 590-8176

© 2021 by The American Association for the Advancement of Science.
All rights reserved.
19 November 2021

### Introduction

**2** On the shoulders of giants
Sean Sanders
*Science*/AAAS

### Foreword

**3** The future is now: The accelerating transformation of medicine
Dennis S. Charney, Eric J. Nestler
Icahn School of Medicine at Mount Sinai

### Artificial Intelligence and Learning Health Systems

**4** The advent of artificial intelligence in health care and its practical implementation
Thomas J. Fuchs, Patricia Kovatch, Emilia Bagiella *et al.*

**5** On the evolution of learning health systems
Alexander W. Charney, Girish N. Nadkarni, Robert Freeman *et al.*

### Biomedical Education

**7** The future of medical, biomedical, and graduate medical education: We will reap what we sow
Marta Filizola, I. Michael Leitman, David Muller

### Cancer

**9** The fight against cancer: Looking back, looking ahead – 50 years of progress
Nina Bhardwaj, Carlos Cordon Cardo, Ramon Parsons *et al.*

### Cardiovascular Medicine

**12** Enhancing cardiovascular health throughout the lifespan: A critical need
Valentin Fuster, Filip K. Swirski, Zahi A. Fayad

### Exposomics

**14** Integrating exposomics into precision medicine and public health
Robert O. Wright, Kecia N. Carroll, Rosalind J. Wright

### Genomic Medicine

**16** Genetics and gene therapeutics: Igniting a revolution in health care
Eimear E. Kenny, Noura S. Abul-Husn, Brian D. Brown *et al.*

### Geriatrics and Palliative Care

**19** Creating new models of care in geriatric and palliative medicine
R. Sean Morrison, Diane E. Meier, Albert L. Siu

### Health Equity

**20** Embedding health equity in biomedical research
Lynne D. Richardson, Gary C. Butts, Carol R. Horowitz

### Immunology

**21** Building the future of immunology in medicine
Miriam Merad, Jeremiah Faith, Dusan Bogunovic *et al.*

### Maternal and Child Health

**24** The frontiers of medical research: Child and maternal health
Siobhan M. Dolan, Lisa M. Satlin, Rachel Vreeman *et al.*

### Microbiology

**26** Preparing for the next pandemic
Florian Krammer, Viviana Simon, Adolfo García-Sastre *et al.*

### Neuroscience

**27** The human brain: Medicine's last frontier
Eric J. Nestler, Yasmin L. Hurd, Paul J. Kenny, *et al.*

### Personalized Medicine

**31** The future of therapeutic discovery: Pioneering new pathways to personalized medicine
Paul J. Kenny, Nina Bhardwaj, Marta Filizola *et al.*

### Stem Cell Biology and Regenerative Medicine

**33** Fulfilling the promise of stem cell biology and regenerative medicine
Joel W. Blanchard , Timothy A. Blenkinsop, Saghi Ghaffari *et al.*

### Surgery

**35** Dramatic innovations in surgery
Joshua Bederson, David H. Adams, Evan Flatow *et al.*

### Women's Health

**39** Comprehensive approaches to sex and gender science
Leslee J. Shaw, Veerle Bergink, Joanne L. Stone *et al.*

# On the shoulders of giants

Medicine—and the science behind it—are advancing quickly. Genomes can be routinely sequenced, certain cancers cured, pandemics attenuated, and the molecular underpinnings of individual cells understood. From the infrastructure and tools to generate and capitalize on enormous datasets, to the ability to edit genetic material, to a more granular and nuanced understanding of the immune system, we now find in the lab and the clinic what was not that long ago found only in science fiction. If these breakthroughs herald the future of health care, what will that future look like?

One thing is for certain: It will include a substantial use of data analytics and artificial intelligence (AI). Pathologists are now training AI-driven algorithms with digitalized tissues to allow faster, more objective, and more quantitatively accurate diagnoses. AI can also inform long-term strategic decision making to enable the most effective use of resources in a pandemic such as that caused by COVID-19.

Technological advances are also enabling improved screening of potential drugs, the identification of drug targets, and the screening of patients who will most greatly benefit from a particular course of treatment—getting us a few steps closer to truly personalized therapies.

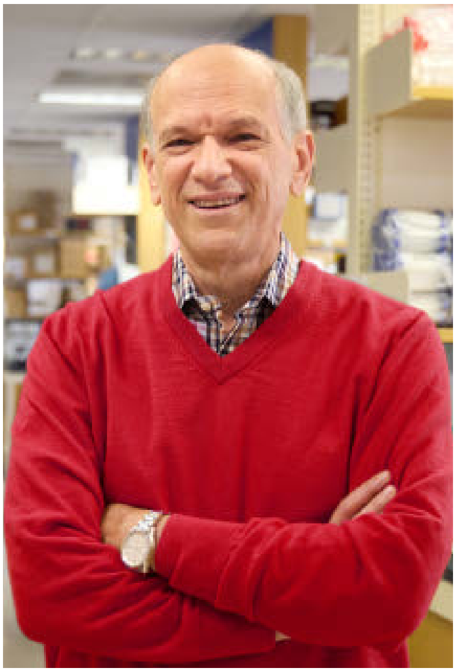
The application of new technologies, together with AI, will allow for earlier, more accurate prediction of adverse side effects, and identification of new and better ways of synthesizing drugs. It may soon allow development of drugs for targets now considered undruggable.

Yet it is not only dramatic, front-page advances that drive medicine and medical research forward. New evidence-based ways of viewing and interacting with patients—including novel screening regimens as well as health-promotion interventions aimed at stemming preventable diseases—can impact health outcomes. We are learning more about how the environment, in conjunction with a patient's genetic makeup, helps determine the biology of health and disease. Researchers are working together to collaborate with each other's intellectual silos, and finding ways to integrate emerging technologies and educational insights into changing how the next generation of practitioners is screened, selected, and mentored, to ensure they are equipped to serve and innovate in accordance with tomorrow's needs and values. This includes awareness of systemic issues that lead to health inequity, a shortened lifespan, and degraded quality of life.

The excitement and optimism around the remarkable progress in medical research and treatment is what inspired the creation of this supplement. The Icahn School of Medicine at Mount Sinai, in partnership with *Science*/AAAS, is spotlighting just some of the extraordinary work being done by researchers and clinicians at the frontiers of medical research. Their hard work and innovative spirit will continue to drive advancements in patient-centered health care, now and into the future.

**Sean Sanders, Ph.D.**
Custom Publishing Office
*Science*/AAAS

Dennis S. Charney



Eric J. Nestler

# The future is now: The accelerating transformation of medicine

This *Science* supplement presents a panorama of the frontiers of biomedical exploration. Never before has this vista been so expansive and exciting as today, as we stand on the cusp of a transformative era of discovery that has the potential to radically advance medicine. The next decade is rich with opportunities for us to recognize disease earlier and with greater precision, to address it more effectively, and ultimately to prevent its occurrence.

A technological revolution has empowered researchers to see deeper into the human body to uncover the etiology and progression of disease. Breakthroughs in DNA sequencing and transcriptomics, for example, are revealing the intricate genetic, epigenetic, and cellular patterns underlying disease, and generating insights that will help solve our most challenging biomedical puzzles and enable the discovery of efficacious new therapies.

In the following 17 brief survey articles, scientists and physicians from the Icahn School of Medicine at Mount Sinai discuss the progress and possibilities within their fields of specialization, including artificial intelligence, cancer, cardiovascular medicine, genomics, immunology, exposomics, and neuroscience.

Today's biomedical frontiers hold great promise. To fulfill this promise, we must continue advancing our scientific understanding and devise novel ways of attacking disease. This requires engaging more diverse and inclusive teams of researchers and clinicians who bring differing perspectives to scientific inquiry. Along with these priorities is the urgency of better harnessing data analytics and artificial intelligence to improve the functioning of hospitals and health care systems.

We envision a world of medicine that is smarter, more equitable and inclusive, and more capable of healing and preventing disease. All this requires more commitment, collaboration, and creativity.

The commitment to science must come in the form of robust research funding from the U.S. National Institutes of Health (NIH), the single most important source of financial support for biomedical research. Measured in inflation-adjusted dollars, NIH funding during fiscal year 2021 was >3% below that of 2003. The following articles stand as evidence in the case for substantially boosting NIH research funding.

The future of medicine depends upon collaboration: active interdisciplinary cooperation across academic institutions and between academia, government, and industry. COVID-19 created an urgency that broke down barriers and enabled the development of assays, new treatments, and vaccines in record time. This historic response to the pandemic can serve as a model of collaboration.

Advancing the frontiers of biomedicine also requires creativity. We need bold thinking from the best and most diverse explorers not only in basic biomedical, translational, and clinical science, but also in biomedical engineering and in physical and computational sciences.

We invite you to join us on the frontiers of biomedicine by reading this inspiring collection of essays that portend sweeping improvements in human health and well-being.

**Dennis S. Charney, M.D.**, is the Anne and Joel Ehrenkranz Dean of the Icahn School of Medicine at Mount Sinai and president for academic affairs for the Mount Sinai Health System.

**Eric J. Nestler, M.D., Ph.D.**, is dean for academic and scientific affairs of the Icahn School of Medicine at Mount Sinai, director of the Friedman Brain Institute, and the Nash Family Professor of Neuroscience.

ARTIFICIAL INTELLIGENCE AND LEARNING HEALTH SYSTEMS

# The advent of artificial intelligence in health care and its practical implementation

**Thomas J. Fuchs[3,4,5]*, Patricia Kovatch[6,7], Emilia Bagiella[2,4], Bruce Darrow[4,8,9], Emma Guttman-Yassky [10], Burton Drayer[11], Zahi A. Fayad[11,12], Carlos Cordon-Cardo[3], Erwin Bottinger[5], Marta Filizola[1,4], David Reich[4,13]**

Our civilization is undergoing a fundamental transformation that will directly impact every single human being. One does not have to agree with Google's CEO that artificial intelligence (AI) is "more profound than electricity or fire" (*1*), but without a doubt, the impact of AI could eclipse advances from the Industrial Revolution, personal computing, and the internet combined. Health care and well-being, as fundamental human needs (*2*), are at the forefront of this transformation.

In health care, this transformation is driven by an annual growth rate of medical data that exceeds all other global data sources, including financial services, manufacturing, media, and entertainment (*3*, *4*). A single digitized microscopy slide, for example, with its millions of cells and morphological structures, contains more data than a full feature high-definition movie. Thus, the advent of AI in health care is not only driven but made necessary by this explosion of data.

Concurrently, the statistical methods used to analyze and model biological and health care data have evolved rapidly, from classical linear models with a few parameters, to early machine learning models such as support vector machines and random forests that depended on manual feature engineering, to state-of-the-art, high-capacity deep-learning approaches with billions of parameters. Examples of the latter in health care include convolutional neural networks for pathology (*5*), dermatology (*6*), radiology (*7*), and drug discovery (*8*), and transformer models for medical language (*9*).

A modern, secure, optimized (*10*, *11*), high-performance computation and data science ecosystem is critical to training such AI systems and realizing their promise. Mount Sinai, for example, has invested significantly in building a comprehensive infrastructure to accelerate research in all biomedical domains based on over 20 petaflops of GPU and CPU compute power, 50 petabytes of storage, and over 100 terabytes of memory. A data commons called "Data Ark" (*12*) contains both public and private datasets, including electronic health record data along with imaging and genomic data. Initiatives like the Data Ark should be guided by the FAIR principles (*13*): making data more findable, accessible, interoperable, and reusable. For example, to help clinicians and researchers make data-driven decisions during the COVID-19 pandemic surge, new datasets were, and still are, made available daily to help track and better understand both individual and collective patient outcomes. An example of a discovery that leveraged this infrastructure was the determination that the origin of the COVID-19 epidemic in New York City arose predominately through untracked transmission between the United States and Europe. The workflows and processes for training such large-scale AI systems as well as their real-world deployment are developed in the new field of MLOps (machine learning operations). Especially useful for clinical deployment in a regulated environment, MLOps is crucial for guaranteeing that AI performs as intended.

For modern health care systems, we are envisioning an "Intelligent Fabric" that underpins all efforts of biomedicine and health care. This Intelligent Fabric will be woven out of AI services that elevate and streamline the work of our nurses, physicians, scientists, and administrators. At the most basic level, we will continue to leverage our existing productivity and decision support tools for diagnosis, treatment, and workflow optimization; and we will connect and interweave these tools together to improve the overall patient journey and optimize the resources for an entire health system. Research analytics and mechanisms for multimodal data integration are not yet built into the modern health care environment. As a result, even health systems that have invested heavily in data science infrastructure remain largely in the dark regarding workflows and outcomes within their walls. At the hospital level, the Intelligent Fabric will overcome this issue by providing insight into the whole system from intertwined AI threads. The aim is to inform local, tactical decision making with information collected from the whole health system. This will empower local actions to be more in tune with other activities taking place throughout the health system, and will also enable long-term strategic decision making for the care of single patients as well as for the entire health care system. The COVID-19 pandemic provides one example of how the Intelligent Fabric is crucial for improving health care delivery: Such a massive, unexpected influx of critically ill patients could easily overwhelm a health care system, necessitating streamlined diagnosis and efficient patient movements between emergency rooms, intensive care units, and other specialized treatment areas. The Intelligent Fabric would greatly assist with system-wide tracking, strategic planning, and proficient execution for situations with large-scale, real-time requirements such as these.

For patients, such an Intelligent Fabric extends beyond improved care within the walls of a health care system: AI companions will also guide them toward disease prevention and healthier living, resulting in enhanced well-being. Intelligent tools will chaperone patients to their appointments and assist with rehabilitation after they return home. The Intelligent Fabric will envelope them with personalized services offering not only the reassuring feeling of being continuously cared for by a health system, but more importantly the confidence that comes with knowing they will get personalized treatment when they are in need of it and will have an active role in the decision-making process.

---

[1]Department of Pharmacological Sciences and Graduate School of Biomedical Sciences; [2]Center for Biostatistics, Department of Population Health Science and Policy; [3]Department of Pathology, Molecular and Cell-Based Medicine; [4]Department of AI and Human Health; [5]Hasso Plattner Institute for Digital Health at Mount Sinai; [6] Department of Genetics and Genomic Sciences; [7]Scientific Computing and Data; [8]Departments of Cardiology, Medicine, and Population Health Science and Policy; [9]Department of Information Technology; [10]Department of Dermatology; [11]Department of Diagnostic, Molecular and Interventional Radiology; [12]BioMedical Engineering and Imaging Institute; [13]Department of Anesthesiology, Perioperative and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: thomas.fuchs.ai@mssm.edu

At Mount Sinai, practical implementation is driven by integrated efforts across the BioMedical Engineering and Imaging Institute (BMEII), the Hasso Plattner Institute for Digital Health at Mount Sinai (HPI-MS), the Charles Bronfman Institute for Personalized Medicine (IPM), and Mount Sinai Centers for Bioinformatics, Biostatistics, and Clinical Intelligence. A new Department of AI and Human Health, the first of its kind located within a medical school in the United States, provides an academic home for AI faculty, enhances Master's and Ph.D. training opportunities (*14*), and facilitates cross-cutting enterprises dedicated to improving patient care through pioneering innovations and technologies.

Developing AI for patient care is a highly challenging task that comes with many responsibilities, ranging from compliance to patient consent, privacy, data protection, and cybersecurity. Wide-ranging AI ethics initiatives and training should ensure that the development, evaluation, validation, and use of AI is trustworthy, and that AI algorithms for clinical care are safe, effective, and equitable. The World Health Organization has published broad guidance on approaching these challenges (*15*). A critical role in assuring this is assigned to regulatory bodies such as the U.S. Food and Drug Administration (FDA) in the United States. While the hype around new technologies is often distracting, in reality, at the time of writing, only a single AI is approved by the FDA for clinical use in pathology.

We hope that the vision laid out in this essay will help modern health care systems and the scientific community in developing, building, and deploying AI in clinical practice, to assure that every patient can benefit from these new and promising technologies.

## References

1. A. Petroff, "Google CEO: AI Is 'More Profound Than Electricity or Fire,'" CNN.com (2018); available at https://money.cnn.com/2018/01/24/technology/sundar-pichai-google-ai-artificial-intelligence/index.html.
2. United Nations Development Programme, "The SDGS in Action" (2021); available at https://www.undp.org/content/undp/en/home/sustainable-development-goals.
3. S. Coughlin, D. Roberts, K. O'Neill, P. Brooks, *Intern. Med. J.* **48**, 92–96 (2018).
4. D. Reinsel, J. Gantz, J. Rydning, "The Digitization of the World from Edge to Core" (Framingham, MA, International Data Corporation, 2018), p. 22; available at http://cloudcode.me/media/1014/idc.pdf.
5. S. Cohen, *Artificial Intelligence and Deep Learning in Pathology* (Amsterdam, Elsevier Health Sciences, 2020).
6. D. H. Murphree *et al.*, *J. Am. Acad. Dermatol.* (2020), doi: 10.1016/j.jaad.2020.05.056.
7. S. K. Zhou *et al.*, *Proc. IEEE* **109**, 820–838 (2021).
8. H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, T. Blaschke, *Drug Discov. Today* **23**, 1241–1250 (2018).
9. Y. Li *et al.*, *Sci. Rep.* **10**, 1–12 (2020).
10. P. Kovatch, L. Gai, H. M. Cho, E. Fluder, D. Jiang, in *Proceedings, 2020 IEEE 34th International Parallel and Distributed Processing Symposium Workshops* (*IPDPSW 2020*), New Orleans, LA, U.S.A., May 18–22, 2020 (IEEE, Piscataway, NJ, 2020), pp. 183–192.
11. P. Kovatch *et al.*, in *SC '15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis,* Austin, TX, U.S.A., November 15–20, 2015 (ACM, New York, 2015), pp. 1–12.
12. Icahn School of Medicine at Mount Sinai, Scientific Computing and Data, "Data Ark: A Data Commons for Mount Sinai" (2021); available at https://labs.icahn.mssm.edu/minervalab/resources/data-ark.
13. M. D. Wilkinson *et al.*, *Sci. Data* **3**, 1–9 (2016).
14. M. Filizola, I. M. Leitman, D. Muller, "The Future of Medical, Biomedical, and Graduate Medical Education: We Will Reap What We Sow," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 7–9.
15. World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance* (World Health Organization, Geneva, 2021).

---

**ARTIFICIAL INTELLIGENCE AND LEARNING HEALTH SYSTEMS**

# On the evolution of learning health systems

**Alexander W. Charney[1]*, Girish N. Nadkarni[1], Robert Freeman[1], Patricia Kovatch[2], Kristin Myers[3], David Reich[4]**

While biomedical research and clinical practice have leaped to new heights, the concept of a learning health system seems stuck in a time warp. Opportunities to fully embed data analytics into daily clinical practice to continuously improve quality of care (*1*) have been missed time and again. Indeed, more than half a century ago, a chief architect of the concept, Guy Octo Barnett, M.D., wrote in the *New England Journal of Medicine*: "The use of computers would, it was held, allow the rapid and accurate collection and retrieval of all clinical information, perform automatic diagnosis, […] interpret all laboratory tests immediately, and replace the telephone and the medical record by fulfilling their functions" (*2*). Five decades later, the contemporary feel of these words is cause for serious contemplation: No true learning health system yet exists. Given this history of underachievement, the guiding principle of Pragmatic Innovation, which we will explore in this essay, can be highly useful in today's health system decision making. First, we touch upon the evolution of learning health systems during this period.

The emergence of the electronic medical record (EMR) as the backbone of health care delivery has been the most critical development in the evolution of learning health systems to date. The predecessors to modern EMRs were computer programs written to streamline specific care processes in specific clinical settings (e.g., the ordering of laboratory tests in a hospital). Their developers—physicians and researchers in academic medical centers—found that a single program to manage all processes was impractical, so they instead designed a "modular system" with different programs for each process (*2*). Over time, the modularity of these systems became highly complex, such that the modern EMR is not simply a log of events as its name might suggest, but rather a virtual space where dozens of independent databases are interconnected to appear to the user as a one-stop shop where tests are ordered, results reported, medications prescribed, and clinicians' thoughts recorded. In parallel to the emergence of EMRs have been efforts to develop clinical decision support tools, which aim to automate diagnosis and predict clinical events. An early example was INTERNIST-1, painstakingly programmed over the course of decades to emulate the ability of a physician to diagnose

---

[1]The Mount Sinai Clinical Intelligence Center, [2]Department of Genetics and Genomic Sciences, [3]Information Technology, Mount Sinai Health System, [4]Department of Anesthesiology, Perioperative and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: alexander.charney@mssm.edu

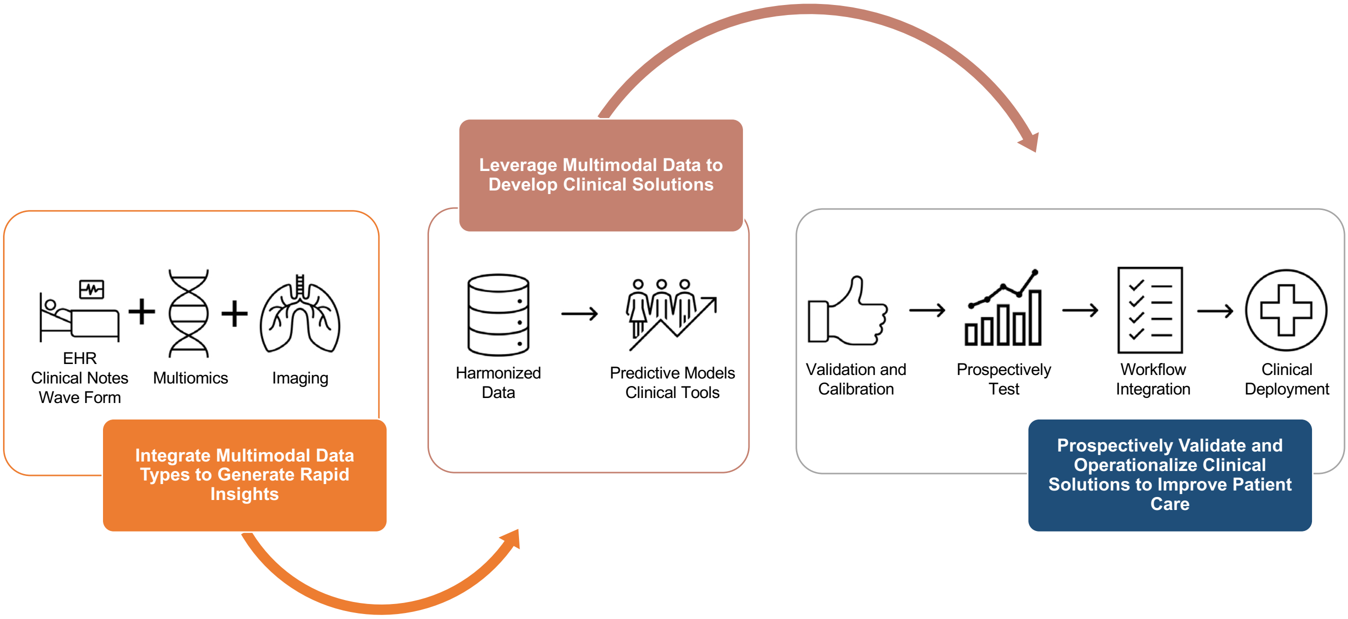

Leverage Multimodal Data to Develop Clinical Solutions
EHR
Clinical Notes
Wave Form
Multiomics
Imaging
Integrate Multimodal Data Types to Generate Rapid Insights
Harmonized Data
Predictive Models
Clinical Tools
Validation and Calibration
Prospectively Test
Workflow Integration
Clinical Deployment
Prospectively Validate and Operationalize Clinical Solutions to Improve Patient Care

**FIGURE 1. A proposed workflow for ensuring the development and deployment of valid and beneficial data science approaches in clinical care.** EHR, electronic health record.

disease (*3*). The earliest iterations of such tools never made it into clinical practice, and their modern derivatives have not had the transformative impact that was promised.

Weaknesses in the digital infrastructure of modern health systems have been laid bare by the COVID-19 pandemic. As frontline providers during this era, we have seen at least four answers to the question Dr. Barnett posed in his time: "What is the cause of this state of underachievement when the need is so great, and the technology supposedly so powerful?" (*2*). First, interconnecting the nodes of modular data in a large health system is both a science and an art, and as a result, these systems are slow to engage. In the years immediately preceding the pandemic, this was a well-known inconvenience, with data once memorably described as "balkanized along organizational boundaries" (*4*). This inconvenience became a missed opportunity of tragic proportions as health systems were strained to the brink of collapse, and as data to guide action could have actually saved lives. Second, even with the nodes connected, research analytics is not built into the modern health care environment. To EMR users, it comes as a surprise to learn that the data they see on a single screen in the EMR resides in dozens of databases on the backend that are not readily accessible for researchers who want to investigate the efficacy of care delivery at the population level. Third, most predictive and other artificial intelligence and machine learning algorithms are not properly validated before being deployed, which may lead to more harm than benefit for the patients and result in physicians developing the habit of ignoring such tools as at best a useless annoyance, and at worst a dangerous detriment to care delivery (*5*). Finally, best practices for dealing with ethical considerations and cybersecurity are often not in place at baseline. Under high-stakes situations where time is of the essence, careful consideration of these complex topics delays decisive action.

In trying to meet the above challenges, we have found that the guiding principle of Pragmatic Innovation pays rich dividends. Specifically, we strive to apply this principle when making decisions in the following key areas:

*Infrastructure*. A modern, secure, high-performance computational and data science ecosystem is critical in the evolution of learning health systems. The solution will differ according to the needs of each institution, and local experts should be identified and tasked with building the requisite infrastructure on premise or buying it in the cloud, depending on the appropriate context. There is no one-size-fits-all solution.

*Centralizing talent*. In the modern health care environment, too often lines are drawn between information technology and data science, quality improvement and research analytics, and data engineers and clinical informaticists. In contrast, every team member in the groups that reshaped health care around EMRs had the same title: "problem solver" (*2*). A consequence of the current organizational structure is a systematic deprioritizing of activities that fall under the umbrella of research analytics, which are the activities most necessary for learning health systems to evolve. These dividing lines between subtly distinct data-centered teams within health care organizations must be replaced with centers of "clinical intelligence" that bring all of the relevant talent under one umbrella.

*Stakeholder engagement and education*. There are many stakeholders in the evolution of learning health systems. These include health care providers, health system leadership, and patients. As we continue to build data-driven clinical tools using advanced analytic techniques such as artificial intelligence, their input must be sought. Most physicians view EMR workflows as an unwanted addition to their administrative burden (*6*). For adoption of data-driven tools, therefore, an upfront investment is needed to educate clinicians on the basic underlying science and to demonstrate to them the concrete evidence of their value

(*7*). Diverse and representative patient advisory groups should also have a voice at all steps in this evolution.

*Prospective validation and randomized testing*. Data science approaches that are not validated or have built-in biases have the potential to cause significant harm when deployed in clinical settings (*8*). Appropriate governance is therefore key, with the inclusion of ethics experts to ensure that patient welfare is always paramount. Any models built should be prospectively validated to ensure no bias, and if bias is detected, it should be corrected. Models and clinical tools should be tested in a randomized, controlled fashion to ensure benefit, using the same rigorous standard currently in place when evaluating an experimental pharmacological intervention. These steps are presented in Figure 1.

In this essay, we have presented achievements to date, challenges ahead, and a roadmap for realizing the complete evolution of learning health systems. With the requisite knowledge, resources, and technology at our disposal, and respect for the difficulty of the task at hand, let's get to work.

## References

1. L. M. Etheredge, *Health Aff. (Millwood)* **26**, w107–w118 (2007).
2. G. O. Barnett, *N. Engl. J. Med.* **279**, 1321–1327 (1968).
3. R. A. Miller, H. E. Pople, J. D. Myers, *N. Engl. J. Med.* **307**, 468–476 (1982).
4. T. Panch, H. Mattie, L. A. Celi, *NPJ Digit. Med.* **2**, 77 (2019).
5. A. Wong *et al.*, *JAMA Intern. Med.* **181**, 1065–1070 (2021).
6. E. W. Jamoom, D. Heisey-Grove, N. Yang, P. Scanlon, *J. Health Med. Inform.* **7**, 1000240 (2016).
7. L. A. Celi *et al.*, *J. Med. Internet Res.* **18**, e325 (2016).
8. J. Zou, L. Schiebinger, *Nature* **559**, 324–326 (2018).

---

**BIOMEDICAL EDUCATION**

# The future of medical, biomedical, and graduate medical education: We will reap what we sow

**Marta Filizola[1*], I. Michael Leitman[2*], David Muller[2*]**

More than 110 years after the Abraham Flexner report (*1*), medical, biomedical, and graduate medical education continue to evolve. Yet a transformational change of the methodologies used generations ago has not occurred, despite the disruption created by the flood of new scientific information and the ability to harness massive amounts of data produced by increasingly complex health care systems. As the proper training of scientists and physicians has profound implications for the future of our species and planet, what we teach, how we teach, and who does the teaching must be fundamentally reimagined, not incrementally tweaked. To enable future breakthrough discoveries that can be rapidly translated into clinical practice and boundless medical and social benefits, we must integrate new scientific knowledge and technological advances into education in a manner that is seamless, cognitively efficient, and values the dignity of every human life.

Despite their differences, the foundation of Ph.D., Master's, and M.D. degrees, and board certification following residency or fellowship, should be governed by core principles that include a rigorous, forward-thinking science education that is bold and creative, not risk averse; a learning and training environment that is nurturing and encourages growth and independent thought; and an approach to equity and justice that eliminates the racism and disparities in privilege that define much of society.

Table 1 describes some of the needs and challenges of science instruction, and some proposed action plans. For instance, the wealth of scientific information produced by advanced genomic, biochemical, and biophysical technologies risks causing cognitive overload if taught using classic faculty-driven, classroom-based teaching modalities. Studies have argued in favor of replacing traditional lecturing with active learning to maximize retention and performance (*2*). The way information is presented also appears to play an important role in the immediate translation of complex information into working memory, as demonstrated, for example, by the positive impact of medical holography in basic anatomy training (*3*). High-definition, three-dimensional virtual reality (VR)/augmented reality (AR) environments and devices, as well as simulators assisted by artificial intelligence (AI)/machine learning tools and algorithms, are revolutionizing science and clinical practice (*4*) and can have the same impact on education, as seen in surgical training (*5*), social cognition training, brain damage assessment and rehabilitation, pain management, and drug discovery (*6*). Challenges associated with the use of VR and AR for both basic research and education include the paucity of content and the difficulty of translating observations into quantitative information. Simulation, on the other hand, has already demonstrated improved accuracy, efficiency, and speed of technical skills in operative outcomes, and has enhanced teamwork, especially related to high-risk but rare events such

---

[1]Department of Pharmacological Sciences and Graduate School of Biomedical Sciences, [2]Department of Medical Education, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding authors: marta.filizola@mssm.edu (M.F); michael.leitman@mssm.edu (M.L.); david.muller@mssm.edu (D.M.)

| Challenges/Needs | Action Plans |
|---|---|
| Fluidity of medical and scientific discoveries.<br>Expanding volume of scientific information produced by advanced genomic, biochemical, biophysical, and cellular technologies, alongside engineering, simulation, and computing principles. | Multimodal instruction that supports flexible learning and teaching styles, including distance learning, team-based platforms, virtual and augmented reality, and simulation. |
| Lifelong learning for trainees, teachers, and mentors. | Ongoing faculty development for an increasingly diverse community of educators, advisors, and mentors. |
| Deeply understand a specialized field while at the same time appreciating the value of all specialties to maximize collaboration. | Leadership development skills that teach team effectiveness and communication, emotional intelligence, resilience, cultural diversity, time management, conflict management, motivating others, managing change, and data analytics for driving quality improvements (*14*). |
| Understanding and assessing the impact, validity, and applicability of artificial intelligence/machine learning tools and algorithms. | First-in-class coursework, training, and research opportunities in artificial intelligence/machine learning, computing, big data analysis and visualization, nanomedicine, genomic sequencing, medical imaging, as well as design and engineering of simulators, medical devices, sensors, and wearable devices.<br>Foundational computational skill sets and deep quantitative training for trainees who are determined to unearth the strengths and limitations of algorithms for advanced analytics.<br>Access to transformational technologies, including simulators, robotics, and genomic sequencing (*16*). |
| Annotate, integrate, and share data from multiple sources. | Ethical reasoning skills to evaluate the dangers of embedding social, political, and economic dimensions into health care decisions, and avoid amplification of health inequities. |
| Developing innovation-minded, entrepreneurial graduates who are critical and creative thinkers, prepared for job opportunities outside as well as inside of academia, have strong science communication skills, and have a penchant for managing academic health centers and shaping policies. | Expert training in communication, leadership, commercialization, intellectual property, and regulatory management. |
| Producing scientists and physicians who are representative of our nation's racial and ethnic demographics. | Early pathways to provide education and mentorship of young learners who are from underrepresented groups in science and medicine. |

**TABLE 1. The future of biomedical science instruction.**

as operating-room fires or intraoperative cardiac arrest (*7*). We must continue to investigate how and why certain methods of instruction augment cognitive performance, thus evaluating their impact and justifying their further development.

Understandably, further development of robotic simulators and other emerging technologies in clinical medicine (e.g., deep learning capabilities, 5G-enabled devices, tricorders, and wearables, among others) will require both educators and learners with quantitative undergraduate degrees and solid computational skills, but also an entrepreneurial attitude and a deep understanding of the culture, missions, and infrastructure of academic health centers and complex health systems. As individuals have different talents, the solution cannot be found in providing foundational training in a variety of different disciplines to a single individual, especially in a medical center where teaching is at the graduate rather than undergraduate level. For example, at Mount Sinai, we have a long history of building interdisciplinary teams by establishing alternative pathways to careers in medicine and biomedical sciences that, despite bypassing many of the traditional premedical (*8*) and biomedical science requirements, continue to produce outstanding graduates from diverse backgrounds. This includes the Humanities and Medicine Program, which was established in 1987; FlexMed (*9*), which expanded this early assurance pathway beyond humanities majors; and the recent Ph.D. concentration in AI and emerging technologies, which exclusively recruits students with quantitative and technical backgrounds. We believe that progress in biomedicine is contingent upon nurturing collegial environments of experts across numerous disciplines, from computational fields to the humanities, as well as the ability to attract and retain domain experts who are in high demand in other industries.

## Roots, sustenance, and soil

If transformational change in education constitutes the roots of this endeavor, and education research is its sustenance, then the learning environment is the soil that must provide the rich substrate for preparing graduates whose careers will span the 21st century. Whom we teach and train will, over time, drive the most profound changes in medicine and biomedical research, as the coming generations of scientists and physicians replace us and refine our values and core principles.

Profound changes in the educational environment have increasingly focused on the experience of learners and their desire to impact society beyond the immediate benefits of scientific discovery, and the diagnosis and treatment of diseases. In recent years, the push for holistic review of candidates that looks beyond GRE, MCAT, and USMLE scores has slowly attracted a more diverse cohort of candidates to medical and graduate schools (*8*, *10*). Following research-driven recommendations (*11*), the Graduate School of Biomedical Sciences at Mount Sinai dropped the GRE score requirement for Ph.D. applications in 2018 and continues to track training outcomes and admissions practices to guide the holistic evaluation of candidates. Given the potential of extracting previously unexplored attributes of candidates or application requirements that are better predictors of training outcomes by data mining with AI-digitized information from candidates' applications, screenings, and interview processes, we will continue to improve admission practices for all of our programs. Regardless of the approach, we are committed to identifying and eliminating biases based on resources, privilege, and social capital, as well as to recruiting student and trainee cohorts who reflect our national race, gender, and socioeconomic demographics (*12*).

As learners enter the field from increasingly diverse and disadvantaged backgrounds, they will bring with them new ways of thinking about health and illness, new approaches to solving problems, and new strategies for sharing, collaborating, and disseminating their work. They will steer science and medicine more intently toward addressing the social determinants, social risk factors, and social needs of patients and populations. As educators, we have an obligation to provide the best educational experience to all of our trainees, which includes attention to their well-being. This may require offering and evaluating training in time management, communication, leadership, resilience, cultural diversity, and overcoming imposter syndrome. Similarly, we must continue to investigate and provide training in the necessary skills today's mentors must possess to address cultural diversity within research-mentoring relationships (*13*, *14*).

Our contribution to the rich substrate of academic health centers includes accelerating the shift away from a culture of bullying and mistreatment that has long been accepted as the norm in academic medicine (*15*), a culture that has tolerated interpersonal and academic behaviors that negatively impact creativity, productivity, camaraderie, and the joy of our work. Most ubiquitous are racism and gender bias, and the outcome has been a crisis of burnout, depression, anxiety, and attrition in our ranks. The cycle of negative behaviors that current faculty experienced from their teachers, mentors, and peers during their own professional development can be broken by establishing an environment that encourages reporting, embraces restorative justice, and confronts head-on the atmosphere of fear of retaliation, real and perceived, that permeates our professional lives. This will yield enormous returns in scientific breakthroughs driven by diversity of thought and an approach to mentorship that is mutually beneficial and enlightening.

Promoting a learning environment that rewards excellence in teaching and mentorship, positive role modeling, and equity, instead of hierarchy, tradition, "paying-your-dues," and abuse, will enable a new generation of empowered physicians and scientists to make the most of what the current scientific revolution has to offer.

## References

1. A. Flexner, *Medical Education in the United States and Canada* (Science and Health Publications, Inc., Washington, DC, 1910).
2. S. Freeman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8410–8415 (2014).
3. M. Hackett, "Medical Holography for Basic Anatomy Training," in *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2013,* Orlando, FL, U.S.A., December 2–5, 2013 (NTSA, Arlington, VA, 2013), Paper No. 13207, pp. 1–10.
4. T. J. Fuchs *et al.*, "The Advent of Artificial Intelligence in Health Care and Its Practical Implementation," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 4–5.
5. J. Bederson *et al.*, "Dramatic Innovations in Surgery," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 35–38.
6. P. J. Kenny, N. Bhardwaj, M. Filizola, A. Stewart, "The Future of Therapeutic Discovery: Pioneering New Pathways to Personalized Medicine," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 31–33.
7. B. Yanagawa *et al.*, *Curr. Opin. Cardiol.* **34**, 571–577 (2019).
8. D. Muller, N. Kase, *Acad. Med.* **85**, 1378–1383 (2010).
9. D. Muller, *Virtual Mentor* **16**, 614–617 (2014).
10. P. Gliatto, I. M. Leitman, D. Muller, *Acad. Med.* **91**, 1498–1500 (2016).
11. J. D. Hall, A. B. O'Connell, J. G. Cook, *PLOS One* **12**, e0169121 (2017).
12. L. Amirhajlou *et al.*, *J. Educ. Health Promot.* **8**, 108 (2019).
13. A. Byars-Winston *et al.*, *J. Clin. Transl. Sci.* **2**, 86–94 (2018).
14. V. Y. Womack *et al.*, *PLOS One* **15**, e0236983 (2020).
15. M. D. Lall *et al.*, *JAMA Netw. Open* **4**, e2121706 (2021).
16. A. Winkler-Schwartz *et al.*, *J. Surg. Educ.* **76**, 1681–1690 (2019).

---

**CANCER**

# The fight against cancer: Looking back, looking ahead—50 years of progress

**Nina Bhardwaj[1,2], Carlos Cordon Cardo[1,3], Ramon Parsons[1,4]*, Joseph Sparano[1,2]**

It has been 50 years since the National Cancer Act was signed into law in 1971, which President Richard Nixon described as the U.S. commitment to a "war on cancer." Over the ensuing decades, this infusion of funding has led to an improved understanding of the causes and biology of many cancers in the laboratory, and to better prevention, screening, and treatment in the clinic. Cancer mortality rates have decreased by nearly one-third since 1971, providing a tangible return on that investment (*1*). Although much of that decline has been due to reductions in smoking and more widespread and earlier cancer screening, improvements in treatment have also contributed to the long-term declines in death from lung, breast, prostate, and colorectal cancer, which are the four leading cancers in the United States.

## New drugs, new options

Given the growing importance of further improving cancer therapeutics, clinical trials have played a critical role in improving outcomes. Such trials increasingly translate basic science discoveries made in the laboratory into tangible benefits in the clinic. For example, 410 novel cancer drugs were approved by the Center for Drug Evaluation and Research at the U.S. Food and Drug Administration over the past decade, including 53 in 2020 alone (*2*).

Entirely new classes of agents have emerged as a direct result of basic science discoveries and emerging technologies, including antibody–drug conjugates, immune checkpoint inhibitors, bispecific T-cell engaging antibodies, and engineered cellular therapy, to name a few. Concurrent with the plethora of new cancer therapeutics has been the rapid evolution of diagnostic molecular tests that identify which individuals benefit

---

[1]Tisch Cancer Institute at Mount Sinai, [2]Department of Medicine, Division of Hematology and Oncology, [3]Department of Pathology, [4]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: ramon.parsons@mssm.edu

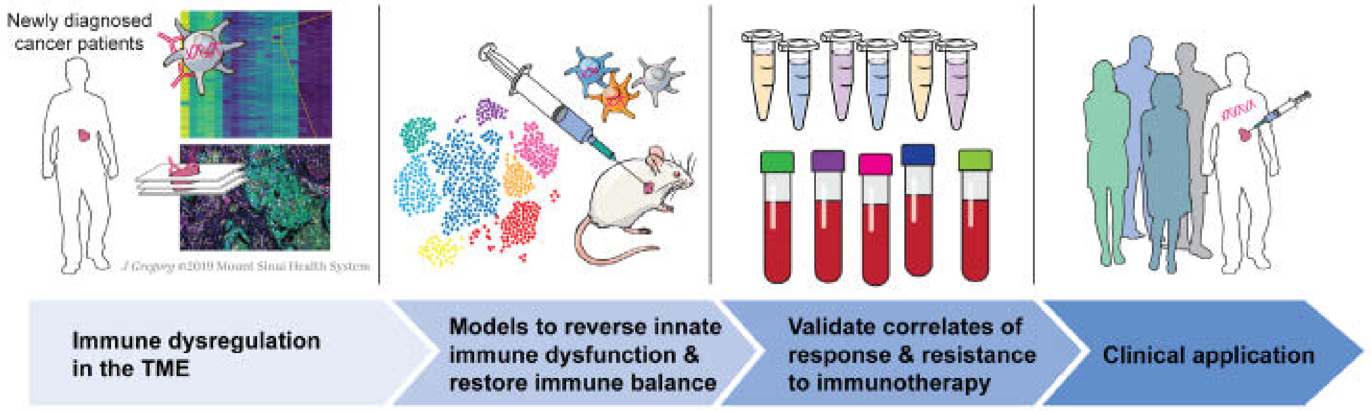

Newly diagnosed
cancer patients
J Gregory ©2019 Mount Sinai Health System
Immune dysregulation in the TME
Models to reverse innate immune dysfunction & restore immune balance
Validate correlates of response & resistance to immunotherapy
Clinical application

**FIGURE 1. Pipeline for the analysis of the tumor microenvironment (TME) to develop clinical applications for patients.**

most from specific therapies, facilitating ever-greater therapeutic precision. The progress in cancer therapy and diagnostics is expected to continue and even accelerate with the development of new technologies.

Although therapeutic clinical trials leading to regulatory approval are generally driven by the pharmaceutical industry, federally funded efforts sponsored by the National Cancer Institute (NCI) have also made major contributions to reducing morbidity and mortality from cancer. NCI supports a robust basic science and clinical research network through its 71 NCI-designated cancer centers and seven NCI-funded research bases that plan and coordinate cancer clinical trials at over 1,000 academic and community sites participating in the National Clinical Trials Network (NCTN) (*3*).

NCI-sponsored trials often address questions that would never be tackled by the pharmaceutical industry, bringing added value. For example, two NCI-sponsored trials in early stage breast cancer, called TAILORx and RxPonder, determined that the majority of women with the most common type of breast cancer could be spared adjuvant chemotherapy after surgery when guided by a diagnostic assay (*4*, *5*). Mount Sinai has a long tradition of clinical trial research that has led to advances in the care of cancer patients. Recent notable success has been achieved in the area of bladder cancer and multiple myeloma, where Tisch Cancer Institute (TCI) investigators took the lead in development of practice-changing therapy (*6*, *7*).

## The hunt for new targets

Fifty years of cancer research has led to the discovery of oncogenes and tumor suppressor genes and to the recognition that mutations and epigenetic changes in these genes participate in the initiation and progression of cancer not only at the cancer-cell level but also by influencing the surrounding cells and tissue. These cancer genes encode proteins that are potential targets for therapy; however, they currently are or have been considered "undruggable." Despite this challenge, many groups including TCI are actively engaged in the study of cancer pathways such as MYC, EGFR/KRAS/BRAF, TGF-β, p53/MDM2, Cyclin D/CDK4, PI3K/PTEN/AKT/FOXO, SWI/SNF, and WNT/β-catenin. Study of these cancer proteins has led to FDA-approved drugs for some targets such as EGFR, BRAF, and CDK4, and novel drug-like tools and drug combinations for others that show promise in preclinical cancer studies (and sometimes human clinical trials) because of their ability to specifically target and inhibit tumor growth in a predictable way based upon the mutation state of a signaling pathway or the expression of a biomarker. Recent examples from TCI include new small-molecule inhibitors and drug combinations that inhibit BRAF/MEK, PI3K, and MDM2, providing therapeutic benefit in different cancer settings (*8*–*10*). Substantial efforts are underway to explore the epigenetic regulation of gene expression that occurs within chromatin. These include efforts in chromatin remodeling and RNA transcription initiation, elongation, splicing, and discovery of noncoding RNAs that modulate tumor growth (*11*–*14*). Finally, the field is working to understand tumor dormancy and microenvironmental signals that activate and suppress the growth of metastatic cells (*15*).

Targeting tumor-cell driver genes that encode protein kinases has led to therapeutic benefit for some cancers through the use of kinase inhibitors, as has the development of DNA repair inhibitors for the treatment of cancers with BRCA mutations. However, new small molecules that can degrade kinases and other proteins that drive cancer growth and metastases will lead to new therapeutic options for previously undruggable targets. Recent highlights from TCI include the development of novel degraders of EGFR, MEK, AKT, EZH2, and CDK4 proteins (*16*–*18*). In addition, replacement of tumor suppressors such as RB, p53, and PTEN has not been feasible in the past, but advances in gene editing should lead to new therapies in this arena as well.

## Improving screening and diagnosis

Fifty years of improvements in the analysis and detection of cancer have also led to much better diagnostic tools and improved cancer screening. The fields of pathology and oncology are embracing cutting-edge, multidimensional diagnostic approaches to classify disease, inform treatment decisions, and train the next generation of academic pathologists. Descriptive anatomic and tissue studies categorizing patients into broad disease types are being refined by more objective molecular analyses aimed at defining individual patient and tumor characteristics. Additionally, the use of cell-based diagnosis is growing, with innovative technologies such as single-cell genotype/phenotype assessments generating highly accurate diagnostics and supporting the now-established cell-based therapeutics for immune modulation. "Digital pathology" is becoming routine, enabling remote expert consultation and more interactive scholarly activities, including multidisciplinary conferences that can now be led by the diagnostician. Digitalized tissue slide banks of images are being used to train and validate artificial intelligence (AI)-driven algorithms, which in turn support the increasing need for objective and quantitative morphometric analysis for more accurate grading and staging of neoplastic diseases (*19*).

In parallel, new developments in clinical pathology and laboratory medicine are being implemented in response to the pursuit of a liquid biopsy platform. This includes the use of mass spectrometry for the measurement of innumerable analytes, such as metabolites and drugs, in a variety of sample types (e.g., serum, urine, and cerebrospinal fluid). A more novel approach involves nuclear magnetic resonance to profile body fluids, encompassing the study of lipid panels, among others.

The rapidly growing field of molecular pathology capitalizes on advances in both tissue and body fluid assessment, aided in large part by microfluidics and nanotechnology. Somatic genetics is becoming the cornerstone of diagnostic and prognostic classification of cancers, both for solid tumors and hematologic malignancies. Profiling of circulating tumor cells and cell-free nucleic acids, along with extracellular vesicles (e.g., exosomes), will bring the next generation of biomarkers.

Pathology is being redefined at the core of patient management and individualized medicine. By integrating complex biochemical, immunological, molecular, and genetic tests, including the systematic application of AI, pathology diagnostics now provide some of the most objective measurements in clinical medicine and are being integrated with image-based methods to improve early detection and measurement of tumor progression. This invaluable information is promoting the development of evidence-based treatment protocols and personalized strategies, optimizing clinical outcomes, and providing cost-effective health care.

Recognition that tumors proactively disable the immune system has led to the development of immunotherapies that have revolutionized cancer treatment through reversal of immune suppression and restoration of immune balance. The effectiveness of checkpoint inhibitors for cancers is a major milestone for the field, however, durable and complete regression has not yet been realized, indicating the persistence of additional pathways of immune dysregulation. We and others are addressing these challenges and expanding the current armamentarium of immune therapies through creation of state-of-the-art facilities to monitor immune function and generate vaccines for cancer patients; comprehensive characterization of mechanisms underlying immune resistance in the tumor microenvironment (TME); establishment of preclinical models to test and validate correlates of resistance; and application of these discoveries into the clinic (Figure 1).

## The importance of biological markers

A current goal for the field is to develop predictive biomarkers for detecting and monitoring graft-versus-host disease that can occur after a bone marrow transplant. These biomarkers hold promise for the better management of this potentially fatal occurrence and therefore ensure better outcomes for cancer patients who require a bone marrow transplant.

Another goal is to create cancer-specific personalized "atlases" that examine the crosstalk between tumor, tumor stroma, and immune cells to characterize immune-resistant TMEs. Through application of functional genomics, proteomics, and cellular characterization, maps of cancers (e.g., bladder, glioblastoma multiforme, lung, and melanoma) have divulged the immune-modulating effects of subclasses of innate and adaptive immune cells. At the end of one spectrum, we have identified macrophage subsets that drive early tumor invasiveness, regulatory T-cell accumulation, and metastasis, and promote deleterious inflammation and angiogenesis (*20*, *21*).

These cellular "biomarkers" have led to novel approaches to control tumor growth through their elimination/modulation in preclinical models and in the clinic (e.g., IL-4 signaling pathway, IL-1 blockade, IRAK4 inhibition, and MERTK inhibitors). At the opposite end of the spectrum, we have demonstrated the critical need for replenishing tumors with specific classes of dendritic cells (DCs), potent antigen-presenting cells essential for cross-presentation of tumor antigens to T cells. By enhancing their mobilization and infiltration into tumors, DCs promote antitumor immunity in humans, further propagated by checkpoint inhibitors and cancer vaccines. Natural killer (NK) cells, another class of innate immune cells, are essential for the recruitment of DCs into tumor sites and for first-line attack against tumor cells. NK cells can become progressively exhausted in the TME; therefore, to promote NK cell function, NK cell-based monoclonal antibody cancer immunotherapies have been developed that inhibit proteolytic cleavage of molecules that enhance tumor cell recognition (MICA A/B, CD16a) (*22*).

Finally, genomic, molecular, and proteomic interrogations have identified new pathways in the effector cytolytic T-cell arm: (1) new immune checkpoints amenable to treatment with monoclonal antibody blockade (e.g., LAG3, TIGIT) or degrader strategies (NKG2A, HPK-1) and (2) off-target tumor-killing pathways (fas-fasL) that can overcome immune escape. Together, these studies are informing the characterization of genomic, molecular, and cellular biomarkers of response to immune intervention in cancer and related disorders.

## Making it personal

Tumor atlases are providing a new appreciation of how a particular individual tumor evades antitumor immunity, emphasizing the need to consider: (1) treating cancers early and (2) personalizing cancer immunotherapy. For example, TCI's Neoadjuvant Tumor Treatment program provides patients who have earlier-stage, potentially resectable cancers the opportunity to receive immunotherapy (usually in the form of checkpoint inhibitors plus another targeted agent) in a short interval prior to tumor resection. Studies are showing the impact of procuring cancer downstaging and even major pathological responses in melanoma, lung cancer, and liver cancer, respectively, setting the stage to broaden this platform of treatment to many cancer types and to identify biomarkers of response. In situ vaccination, namely modulating and activating a tumor site through injection of potent adjuvants or oncolytic viruses, renders a tumor its own "vaccine factory." Investigators have shown the efficacy of such approaches in lymphoma, for example (*23*).

Personalized cancer vaccine development is a major current effort for the field. For instance, the Personalized Genomic Vaccine (PGV) platform at TCI is an outcome of in-house science that uses in-house algorithms to predict patients' expressed personalized genomic neoantigens. Ongoing trials in melanoma, glioblastoma, bladder cancer, multiple solid tumors, and multiple myeloma are showing early evidence of immunogenicity and correlation with clinical outcome in combination with checkpoint inhibitors, tumor treatment fields, or the adjuvant setting. Novel messenger RNA (mRNA) platforms are also being exploited in neoantigen vaccine formulations. The recognition and discovery of novel "shared" frameshift-derived neoantigens in high levels of microsatellite instability (MSI-H) tumors that correlate with response to immunotherapy, and myeloproliferative neoplasms (e.g., frameshifts in the calreticulin gene), respectively, are paving the way for development of personalized yet "shared neoantigen" vaccines. The next chapter for cancer immunology will augment functional genomics platforms to incorporate microbiome and metabolomic analyses to further define an individual's immune response to their cancer, demarcate how it might affect their prognosis, guide the treatment they receive, and predict their response to therapy.

Cancer prevention and early detection are extremely important for reducing cancer mortality, and strides in this area have been made in improving screening for breast, prostate, lung, and colorectal cancer nationally with contributions from TCI in all of these major areas. Notable recent TCI achievements in epidemiology and public health have included linking prostate inflammation to elevated prostate cancer incidence in 9/11/2001 first responders at the World Trade Center, identifying increased levels of insulin resistance among African American women with newly diagnosed breast cancer, and increasing human papillomavirus (HPV) vaccination in teenagers (*24*, *25*). We are excited by all of the progress made in the past and by the momentum toward a future that can fulfill our mission to overcome cancer.

## References

1. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, *CA Cancer J. Clin.* **71**, 7–33 (2021).
2. Statista.com, "Total Number of Novel Drugs Approved by CDER from 2008 to 2020"; available at https://www.statista.com/statistics/817534/annual-novel-drug-approvals-by-cder.
3. National Cancer Institute, "NCI Community Oncology Research Program (NCORP)"; available at https://www.cancer.gov/research/infrastructure/clinical-trials/ncorp.
4. J. A. Sparano *et al.*, *N. Engl. J. Med.* **380**, 2395–2405 (2019).
5. J. Jasem *et al.*, *J. Natl. Compr. Canc. Netw.* **15**, 494–503 (2017).
6. D. F. Bajorin, S. Collette, *N. Engl. J. Med.* **385**, 957 (2021).
7. A. Chari *et al.*, *N. Engl. J. Med.* **381**, 727–738 (2019).
8. C. Adamopoulos *et al.*, *Cancer Discov.* **11**, 1716–1735 (2021).
9. Z. M. Khan *et al.*, *Nature* **588**, 509–514 (2020).
10. J. Mascarenhas *et al.*, *Blood* **134**, 525–533 (2019).
11. Z. A. Qadeer *et al.*, *Cancer Cell* **36**, 512–527.e9 (2019).
12. P. K. Parua *et al.*, *Nature* **558**, 460–464 (2018).
13. J. Y. Fong *et al.*, *Cancer Cell* **36**, 194–209.e9 (2019).
14. R. Rodriguez-Barrueco *et al.*, *Genes Dev.* **31**, 553–566 (2017).
15. J. A. Aguirre-Ghiso, *Science* **361**, 1314–1315 (2018).
16. B. Dale *et al.*, *Nat. Rev. Cancer* **21**, 638–654 (2021).
17. J. Xu *et al.*, *Cancer Discov.* (2021), doi: 10.1158/2159-8290.CD-20-0815.
18. X. Wu *et al.*, *Nat. Cancer* **2**, 429–443 (2021).
19. L. M. da Silva *et al.*, *J. Pathol.* **254**, 147–158 (2021).
20. L. Wang *et al.*, *Clin. Cancer Res.* **27**, 4287–4300 (2021).
21. B. Maier *et al.*, *Nature* **580**, 257–262 (2020).
22. L. Ferrari de Andrade *et al.*, *Science* **359**, 1537–1542 (2018).
23. L. Hammerich *et al.*, *Nat. Med.* **25**, 814–824 (2019).
24. Y. Gong *et al.*, *Mol. Cancer Res.* **17**, 1605–1612 (2019).
25. E. J. Gallagher *et al.*, *Breast Cancer Res.* **22**, 40 (2020).

---

**CARDIOVASCULAR MEDICINE**

# Enhancing cardiovascular health throughout the lifespan: A critical need

**Valentin Fuster[1,2]*, Filip K. Swirski[3], Zahi A. Fayad[4]**

The circulatory—or cardiovascular—system, comprising the heart, blood, lymph, and vessels, carries oxygen, carbon dioxide, nutrients, cells, and their products across all tissues and organs in the body. Serving the body's transportation needs, the heart and vasculature are indispensable to the immune, metabolic, endocrine, hematopoietic, and nervous systems. A healthy life requires a robust and efficient cardiovascular function, yet many environmental and lifestyle factors can jeopardize and weaken it, increasing risk of disease. It is essential therefore to focus on cardiovascular health (CVH) throughout the lifespan, paying particular attention to childhood, middle age, and old age. We are leveraging Mount Sinai's rich history of discovery in cardiology with innovations in science and technology, and are committed to the idea that CVH needs to be understood if it is to be cultivated and enriched.

Throughout life, the processes constituting CVH change as we age. Sleep, for example, strongly associates with heart health, but varies dramatically across the lifespan, generally diminishing in need over time. Comorbidities that accrue throughout our lives, whether acquired genetically or environmentally, will affect the cardiovascular system, and their specific influence

---

[1] Mount Sinai Heart, Mount Sinai Medical Center, and the Department of Medicine (Cardiology), Icahn School of Medicine at Mount Sinai, New York, NY

[2] Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain

[3] Cardiovascular Research Institute, and the Departments of Medicine (Cardiology) and Radiology, Icahn School of Medicine at Mount Sinai, New York, NY

[4] BioMedical Engineering and Imaging Institute, and the Departments of Radiology and Medicine (Cardiology), Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: valentin.fuster@mssm.edu

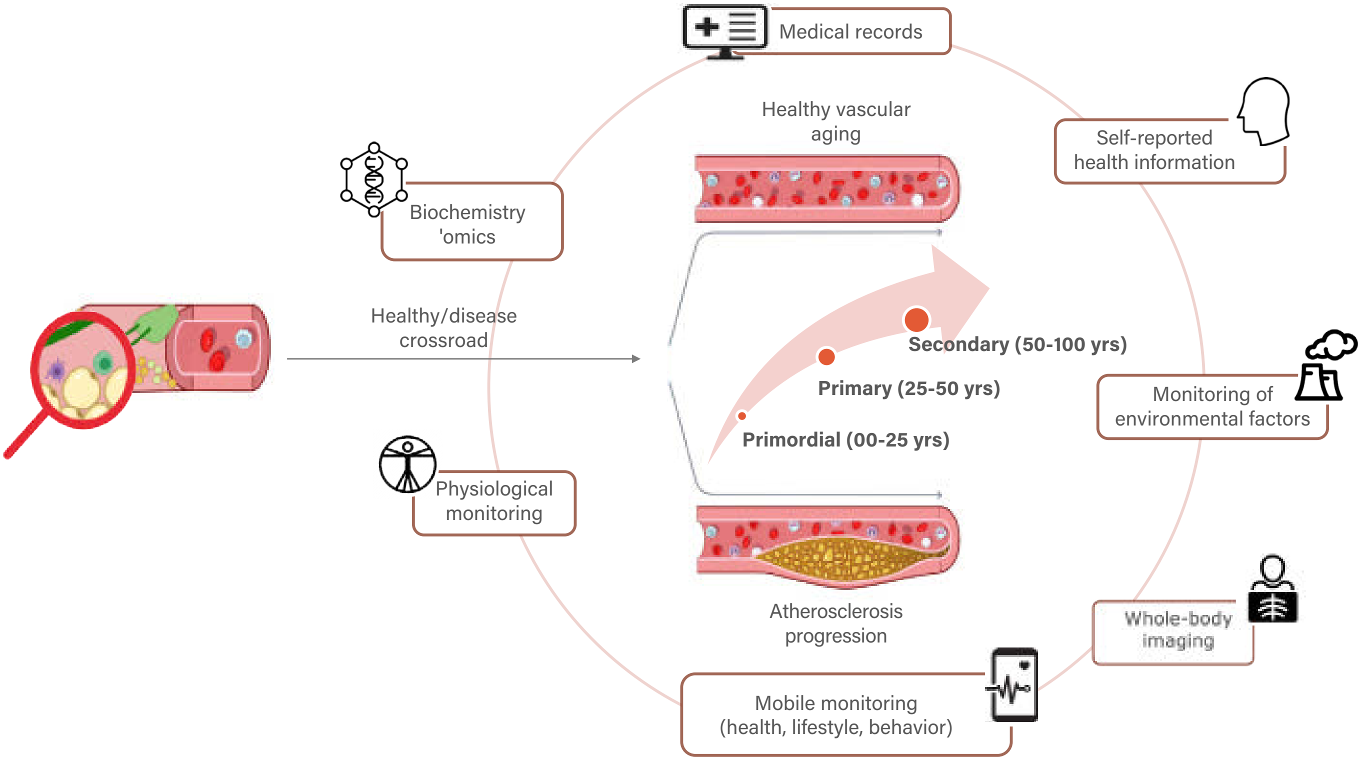

Medical records
Healthy vascular aging
Self-reported health information
Biochemistry 'omics
Healthy/disease crossroad
Secondary (50-100 yrs)
Primary (25-50 yrs)
Primordial (00-25 yrs)
Monitoring of environmental factors
Physiological monitoring
Atherosclerosis progression
Whole-body imaging
Mobile monitoring (health, lifestyle, behavior)

**FIGURE 1. Deep phenotyping of cardiovascular disease throughout the lifespan**

will vary. Some cardiovascular diseases affect younger people disproportionately, while others hardly occur until old age. As aging is a process involving the interplay of environment and genetics as a function of time, it is essential to track the dynamics of CVH separately in children, middle-aged people, and the elderly.

Mount Sinai is working closely with other institutions, including the National Centre for Cardiovascular Research in Madrid (CNIC), to track CVH in these three cohorts. The SI! Program for Cardiovascular Health Promotion in Early Childhood has been applied at all preschool levels in 24 schools, including 2,062 children, ages 3 to 5 years, to promote CVH among children using their proximal environment (school, teachers, and families). Lessons have been learned over 10 years of experience implementing the SI! Program for CVH promotion in preschool settings in three locations: Bogotá (Colombia), Madrid (Spain), and Harlem (New York City, United States). Based on this experience, by matching rigorous efficacy studies with implementation science, we can now help bridge the divide between science and educational practice for nearly 50,000 children. It will require the integration of several factors: (1) multidisciplinary teams, (2) multidimensional educational programs, (3) multilevel interventions, (4) local program coordination and community engagement, and (5) scientific evaluation through randomized studies. The concept behind the program is that implementing effective health promotion interventions early in life may induce long-lasting, healthy behaviors that could help to curb the CVD epidemic (*1*–*4*).

Atherosclerosis starts early in life and progresses silently for decades before manifesting as debilitating or fatal cardiovascular events, a fact that imposes a significant challenge to understanding the disease. Accordingly, early adulthood and middle age are important targets for investigating CVH. Recognizing that atherosclerosis does not simply affect the vessel wall but is also systemic invites the use of noninvasive methods to identify disease before clinical symptoms appear. In examining the effectiveness of adult intervention through primary and secondary prevention, the age range of 25–50 years appears to offer an opportunity to evaluate subclinical disease, about which we have been learning a significant amount through noninvasive imaging techniques. In two recent bioimaging studies, we assessed approximately 10,000 asymptomatic adults age >40 years using multimodality vascular imaging of the carotid and iliofemoral arteries using two- and three-dimensional ultrasound and of the coronary arteries with computed tomography techniques. Focusing on early adulthood and middle age, the multiple lines of clinical and experimental investigation elucidating the biology of CVH require sophisticated tools (*1*). This multidisciplinary science harnesses imaging modalities that have a long history in cardiology, such as cardiovascular magnetic resonance imaging and positron emission tomography (PET), along with the aforementioned computed tomography and ultrasound; in aggregate, these are used to visualize and characterize heart and vessel function as well as their interactions with other tissues and organs such as the brain, body fat, and others (*3*). We utilize classic flow cytometry and microscopy tools and augment them with genomics, proteomics, metabolomics, and other 'omics approaches for highly detailed exploration of the systems' complexities (deep phenotyping; see Figure 1). We are leveraging the continuous monitoring of health and multimodal data collection throughout the lifespan. Specifically, we are securely utilizing sensors, electronic medical records, laboratory tests, imaging data, self-reported information via mobile apps, genetic and molecular tests, and pathology. With artificial intelligence and machine learning, we are building and updating a "digital twin" or "avatar" from every subject throughout their lifespan (*5*, *6*).

A dominant focus in cardiovascular science and clinical cardiology involves elucidating and targeting processes that both precede and follow acute cardiovascular events, which typically, though not exclusively, occur later in life. The rupture or erosion of atherosclerotic plaques may give rise to myocardial infarction or stroke; those who survive such events have elevated risk for reinfarction and complications such as heart failure. Even though a wide array of therapeutics, devices, and surgical interventions are available and highly effective at mitigating risk, chronic cardiovascular disease, and the acute events that it precipitates, continue to claim lives. We must therefore gain a deeper and fundamental understanding of the circulatory system, and to do so, we must learn how the system responds to environmental perturbations and interacts with other organs and tissues (*7*). Current research devoted to CVH and disease connects the influence of the environment and modifiable risk factors with molecular, cellular, and tissue processes throughout the body. For instance, we are exploring how diet, sleep, exercise, and stress affect physiology, studying not only the heart and blood vessels, but also the contribution of critical nervous, immune, metabolic, and endocrine networks (*8*–*11*). Diseases of aging, including neurodegenerative and neuroinflammatory diseases such as Alzheimer's disease, are characterized by similar and convergent molecular pathways known to be important in classical vascular diseases such as atherosclerosis.

A systems biology (*12*) conceptual framework is guiding ongoing research at Mount Sinai and elsewhere, as the field seeks to integrate environmental inputs with interorgan communication that influences health and homeostasis. It is this explorative fusion of harnessing sophisticated tools to examine lifestyle, environment, and biology in the young, middle aged, and elderly, coupled with an increasing emphasis on salutogenesis (which focuses on the origins of health and well-being) (*13*) rather than disease exacerbation, that is poised to transform health care.

## References

1. B. Ibanez *et al.*, *J. Am. Coll. Cardiol.* **78**, 156–179 (2021).
2. J. V. Turco, A. Inal-Veith, V. Fuster, *J. Am. Coll. Cardiol.* **72**, 908–913 (2018).
3. L. Fernández-Friera *et al.*, *J. Am. Coll. Cardiol.* **73**, 1371–1382 (2019).
4. G. Roth, G. Mensah, V. Fuster, *J. Am. Coll. Cardiol.* **76**, 2980–2981 (2020).
5. J. Corral-Acero *et al.*, *Eur. Heart J.* **41**, 4556–4564 (2020).
6. T. J. Fuchs *et al.*, "The Advent of Artificial Intelligence in Health Care and Its Practical Implementation," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 4–5.
7. F. K. Swirski, M. Nahrendorf, *Nat. Rev. Immunol.* **18**, 733–744 (2018).
8. C. S. McAlpine *et al.*, *Nature* **566**, 383–387 (2019).
9. C. J. O'Donnell *et al.*, *JAMA Cardiol.* (2021), doi: 10.1001/jamacardio.2021.2530.
10. M. T. Osborne *et al.*, *Circ. Cardiovasc. Imaging* **13**, e010931 (2020).
11. A. Tawakol *et al.*, *Lancet* **389**, 834–845 (2017).
12. C. Calcagno, W. J. Mulder, M. Nahrendorf, Z. A. Fayad, *Arterioscler. Thromb. Vasc. Biol.* **36**, e1–e8 (2016).
13. H. M. Langevin, *J. Altern. Complement. Med.* **27**, 623–626 (2021).

---

**EXPOSOMICS**

# Integrating exposomics into precision medicine and public health

**Robert O. Wright*, Kecia N. Carroll, Rosalind J. Wright**

Precision medicine, which now primarily utilizes genomic and electronic health record data, holds the potential to transform medical practice. Recent calls to expand the field's purview to encompass *all* factors that predict health (*1*–*3*) underscore the need to include environment (*4*, *5*), as human health and disease are shaped by a range of lifelong exposures and their resultant biological responses. *The emerging exposome concept addresses this complexity by studying the effects of all health-relevant environmental factors over the life course* (*6*–*8*). The 20th-century concept of "nature *versus* nurture" needs to be tossed aside, as genetics and environment do not compete—they work hand in hand through interactions in which our environment (i.e., where people are born, live, develop, learn, work, play, and age) triggers biological responses that are determined in part by our genetics in a lifelong, iterative process. However, there is not yet a systematic plan to integrate exposomics into precision medicine. Advances in computational science, medical informatics, remote sensing, geographical information systems, and analytical chemistry now provide powerful research tools to facilitate the incorporation of exposomics into precision medicine and public health (*9*–*11*).

Understanding our environment is critical to advancing human health and eliminating health inequities. The broad-ranging effects of climate change consequent to rising global average temperatures, increasing atmospheric carbon dioxide ($CO_2$), associated extreme weather events (i.e., heat waves, droughts, floods, and hurricanes), and altered ecosystem functioning and biodiversity, have alarming potential to magnify toxic environmental exposures and amplify consequent health disparities (Figure 1) (*12*). Climate affects the exposome both directly and indirectly to influence the spatial and temporal distribution of infections, asthma, allergies, mental health, and dozens of other chronic diseases. Also, rates of cancer, diabetes, heart disease, and neurodegenerative diseases, among other complex diseases, continue to rise with minority groups such as Blacks, Latinx, Native Americans, and Asian Americans experiencing increased disease morbidity and/or severity. Structural racism, including redlining, has segregated

---

Department of Environmental Medicine and Public Health and the Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: robert.wright@mssm.edu

minoritized groups to live in low-income communities where increased exposure to psychosocial stressors, air pollution, physical environmental toxicants, and reduced access to healthy food and green spaces are concentrated. The systematic segregation of persons of color through redlining policies has significant public health implications (*13*). These environmental disparities go beyond pollution, particularly for indigenous populations, as cultural practices commonly involve local food and water, further exacerbating social inequities. Health inequities arise from variable—and too often predictable—imposed environments that link to social determinants of health. The geospatial diversity of the exposome underlies its key role in explaining health inequities. Broad categories of the human exposome include nutrition, infections, chemicals, physical environment, and sociocultural conditions that act as psychosocial stressors. Because exposomics encompasses the interconnections among our social, nutritional, and chemical environments, it holds great promise for elucidating root causes of recalcitrant health disparities.

The technology to measure the exposome is far more advanced than many researchers may realize, given that it has arisen largely in parallel silos. At Mount Sinai, our Institute for Exposomic Research brings together the broad intellectual and technological expertise that moves us toward unifying exposomics into a single, cohesive field. For example, we now utilize untargeted chemical assays that assess hundreds of thousands of endogenous and exogenous small molecules in blood or urine (*14*). We have also developed novel methods using nontraditional biological matrices (teeth, hair) that reconstruct our distant past, even fetal life, giving us the time-varying data on environment and other biological markers needed to understand the developmental origin of health and disease (*15–18*). At the other end of the spectrum, we measure the external environment using satellites and other remote sensors that record our daily climate, air pollution, noise, green space, and built environment. Electronic health records are increasingly providing not only phenotypic data, but also environmental data, including occupation, address, drugs prescribed, medical procedures, and dates of hospitalizations. Our institute complements these measures with state-of-the-art computational infrastructure to enable the mining of public and administrative databases to estimate chemical exposures, infections (*19–23*), socioeconomic status, urban food deserts, and neighborhood crime levels. Cellphones, wearable devices, and the Internet of Things (IoT) are ubiquitous and can link an individual's place in space and time to concurrent environmental conditions and physiologic reactions. All of these methods, many developed only in the last decade, can now be brought together to define the exposome across our lifespan. Measuring the time-varying nature of exposomics and the societal context that shapes an individual's cumulative exposures may seem daunting, but provides unique opportunities for life-course research. Work is needed to integrate all of these measures into large diverse populations in order to conduct multiomic research on complex diseases, with the exposome being the layer at which many of the interventions will occur.

Exposomics is at a developmental inflection point, as the realization of its commercial, public health, and medical value is only beginning to take hold. We are now calling for a reference exposome study analogous to the genomic Hapmap Project that characterized haplotypes in populations around the world. Exposomic "exposotypes" will vary by sex, culture, geography, calendar year, and life stage (e.g., children will be different from adults). A reference exposome will enable comparisons across these variables and will be enormously beneficial to researchers while enabling better causal inference analyses. The exposome must also become part of precision medicine initiatives. The global population is getting sicker, and

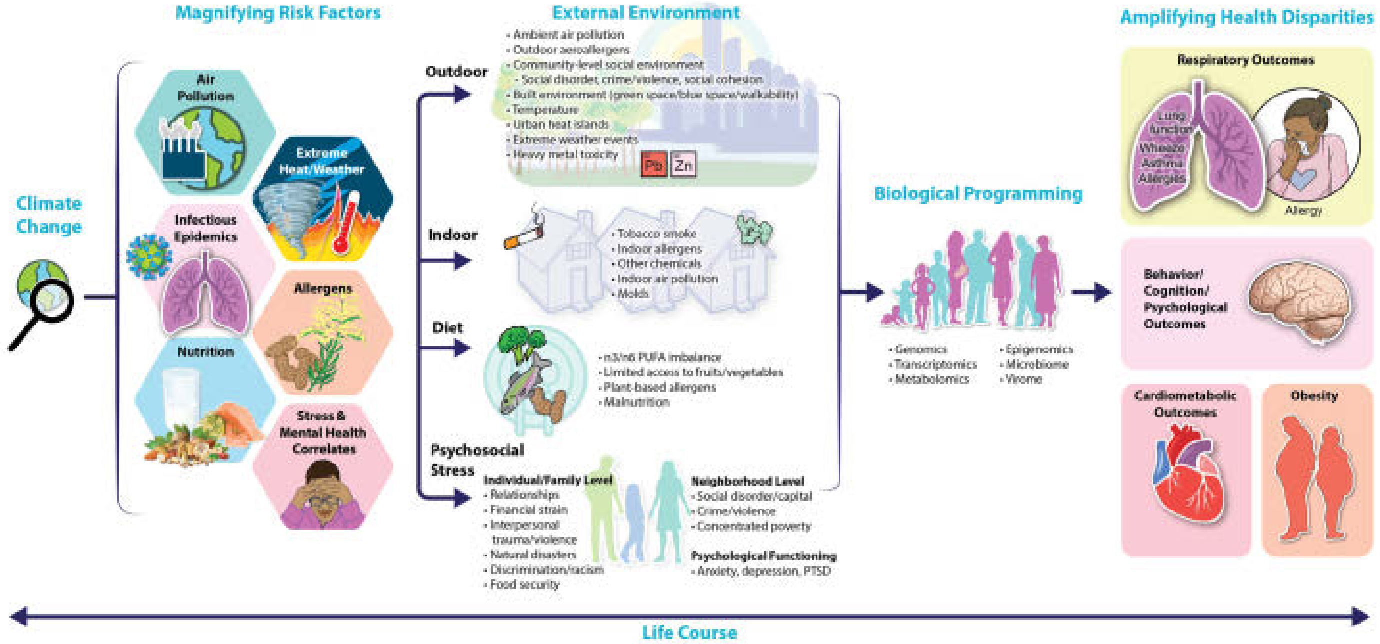

Magnifying Risk Factors
Climate Change
Air Pollution
Extreme Heat/Weather
Infectious Epidemics
Allergens
Nutrition
Stress & Mental Health Correlates
External Environment
Outdoor
Ambient air pollution
Outdoor aeroallergens
Community-level social environment
Social disorder, crime/violence, social cohesion
Built environment (green space/blue space/walkability)
Temperature
Urban heat islands
Extreme weather events
Heavy metal toxicity
Pb
Zn
Indoor
Tobacco smoke
Indoor allergens
Other chemicals
Indoor air pollution
Molds
Diet
n3/n6 PUFA imbalance
Limited access to fruits/vegetables
Plant-based allergens
Malnutrition
Psychosocial Stress
Individual/Family Level
Relationships
Financial strain
Interpersonal trauma/violence
Natural disasters
Discrimination/racism
Food security
Neighborhood Level
Social disorder/capital
Crime/violence
Concentrated poverty
Psychological Functioning
Anxiety, depression, PTSD
Biological Programming
Genomics
Transcriptomics
Metabolomics
Epigenomics
Microbiome
Virome
Amplifying Health Disparities
Respiratory Outcomes
Lung function
Wheeze
Asthma
Allergies
Allergy
Behavior/ Cognition/ Psychological Outcomes
Cardiometabolic Outcomes
Obesity
Life Course

**FIGURE 1. Role of exposomics as a tool to understand direct and indirect effects of climate change on human health and health disparities.** Climate impacts all aspects of the exposome, and its effects are distributed variably depending on location, influencing urban, suburban, and rural communities differently. Image from J. Gregory ©2021, Mount Sinai Health System.

those with chronic illness comprise an increasing proportion of our population (*24*, *25*). Patients experience environments that may not have caused their disease but can nonetheless *impact* their disease or its treatment—addressing this kind of problem is at the very heart of precision medicine. As an example, we should consider the possibility that chemical obesogens, such as perfluoryl alkyls that are ubiquitous in the U.S. population, are not just a cause of diabetes, but may also modify the response to exogenous insulin and should be studied as a *cause of treatment failures or as an initiator of disease complications*. Diabetic patients who require escalating doses of insulin will benefit from an environmental assessment to determine modifiable risk factors that are driving exacerbation of hyperglycemia. Exposomics will also lend itself to complementary partnerships with genomic and health inequity researchers, as all diseases have an environmental component, just as all diseases have a genetic component. By committing to exposomics, we will drive 21st-century research toward a greater understanding of the interactions between our genome and our exposome.

## References

1. M. M. Davis, T. P. Shanley, *Clin. Transl. Sci.* **10**, 64–66 (2017).
2. G. Gibson, *Adv. Exp. Med. Biol.* **799**, 1–14 (2014).
3. R. I. Horwitz, A. Hayes-Conroy, B. H. Singer, *Psychother. Psychosom.* **86**, 5–10 (2017).
4. A. P. Hills, N. J. Farpour-Lambert, N. M. Byrne, *Prog. Cardiovasc. Dis.* **62**, 34–38 (2019).
5. C. Jiang *et al.*, *Cell* **175**, 277–291.e31 (2018).
6. G. W. Miller, D. P. Jones, *Toxicol. Sci.* **137**, 1–2 (2014).
7. R. O. Wright, *Curr. Opin. Pediatr.* **29**, 211–217 (2017).
8. R. O. Wright, S. Teitelbaum, C. Thompson, D. Balshaw, *Curr. Opin. Pediatr.* **30**, 285–291 (2018).
9. P. D. Juarez *et al.*, *Int. J. Environ. Res. Public Health* **11**, 12866–12895 (2014).
10. T. J. Oyana, P. Matthews-Juarez, S. A. Cormier, X. Xu, P. D. Juarez, *Int. J. Environ. Res. Public Health* **13**, ijerph13010013 (2015).
11. G. M. Buck Louis, M. M. Smarr, C. J. Patel, *Curr. Environ. Health Rep.* **4**, 89–98 (2017).
12. R. J. Wright, *Lancet Child Adolesc. Health* **4**, 859–860 (2020).
13. L. F. Novick, *J. Public Health Manag. Pract.* **27**, 435–436 (2021).
14. D. Valvi *et al.*, *Environ. Int.* **143**, 105957 (2020).
15. C. Austin *et al.*, *Sci. Rep.* **6**, 18802 (2016).
16. C. Figueroa-Romero *et al.*, *Ann. Clin. Transl. Neurol.* **7**, 872–882 (2020).
17. N. Nair *et al.*, *Gastroenterology* **159**, 383–385 (2020).
18. M. Yu *et al.*, *Environ. Int.* **157**, 106849 (2021).
19. D. K. Barupal, S. F. Baygi, R. O. Wright, M. Arora, *Front. Public Health* **9**, 653599 (2021).
20. D. K. Barupal, M. K. Schubauer-Berigan, M. Korenjak, J. Zavadil, K. Z. Guyton, *Environ. Int.* **156**, 106624 (2021).
21. I. Gutiérrez-Avila *et al.*, *Int. J. Climatol.* **41**, 4095–4111 (2021).
22. L. Petrick *et al.*, *Leuk. Res.* **106**, 106585 (2021).
23. K. Schroeder, J. G. Noll, K. A. Henry, S. F. Suglia, D. B. Sarwer, *Prev. Med. Rep.* **23**, 101501 (2021).
24. C. Aguayo-Mazzucato *et al.*, *Diabetes Metab. Res. Rev.* **35**, e3097 (2019).
25. J. M. Lawrence, *JAMA* **326**, 717–727 (2021).

---

**GENOMIC MEDICINE**

# Genetics and gene therapeutics: Igniting a revolution in health care

**Eimear E. Kenny[1*], Noura S. Abul-Husn[1,2], Brian D. Brown[3], Judy H. Cho[4], Alison M. Goate[5]**

## Introduction

Starting with the completion of the Human Genome Project two decades ago (*1*), genetics research has exploded and today plays a central and catalytic role in biological research and health care. Spurred by an over 1-million-fold reduction in the cost of genome sequencing and increasingly sophisticated computational methods, data interrogating the link between genetic variants and human health have been generated and analyzed at a massive scale in both research and clinical settings. This has allowed researchers to determine the genetic basis for thousands of rare and common diseases, and to gain an ever-deepening understanding of our biological complexity. Importantly, this research success has been converted into powerful tools for health care, driving the development of novel diagnostic, preventative, and therapeutic strategies.

Mount Sinai has long been at the forefront of genetics research and genomic medicine. It houses the first department of genetics in an academic medical school in the United States, and has a track record of successful treatment development for heritable diseases, including creating one of the first U.S. Food and Drug Administration (FDA)-approved enzyme replacement therapies—Fabrazyme (*2*). Investments in high-performance computing and large-scale biobanking are enabling a future at the interface of genetics, artificial intelligence, and precision medicine.

## Discovery genetics at a massive scale

A major advance for discovery genetics has been the increasing availability of health system–based biobanks, which offer an unprecedented ability to examine the health impact of disease-associated genes and variants in real-world settings (*3*). The Bio*Me* Biobank, established in 2006, is a large, exceptionally diverse biobank drawing from the multiethnic patient population in New York City and supporting over $50 million in research funding. The rich scope of genetic, demographic, clinical, and lifestyle data in this biobank accelerates the discovery of variants that impact risk for both rare syndromes and common diseases such as obesity, diabetes, and autism (*4*–*6*). The data often connect diseases in surprising ways to reveal shared biology and overlapping mechanisms, such as variants that influence multiple autoimmune conditions (*7*). Research is revealing the genetic complexity of schizophrenia and how genomic information can be used to predict disease risk (*8*), as well as the challenges of translating this information into clinical care (*9*), highlighting the need to refine genetic risk prediction to successfully bring genomics into mental health care delivery. We have learned that host–genome interactions have substantially shaped the architecture of many human diseases, a point that has been vividly underscored by marked interindividual

---

[1]Institute for Genomic Health, Department of Medicine, [2]Division of Genomic Medicine, Department of Medicine, [3]Icahn Genomics Institute, Department of Genetics and Genomic Sciences, [4]Charles Bronfman Institute for Personalized Medicine, Department of Genetics and Genomic Sciences, [5]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: eimear.kenny@mssm.edu

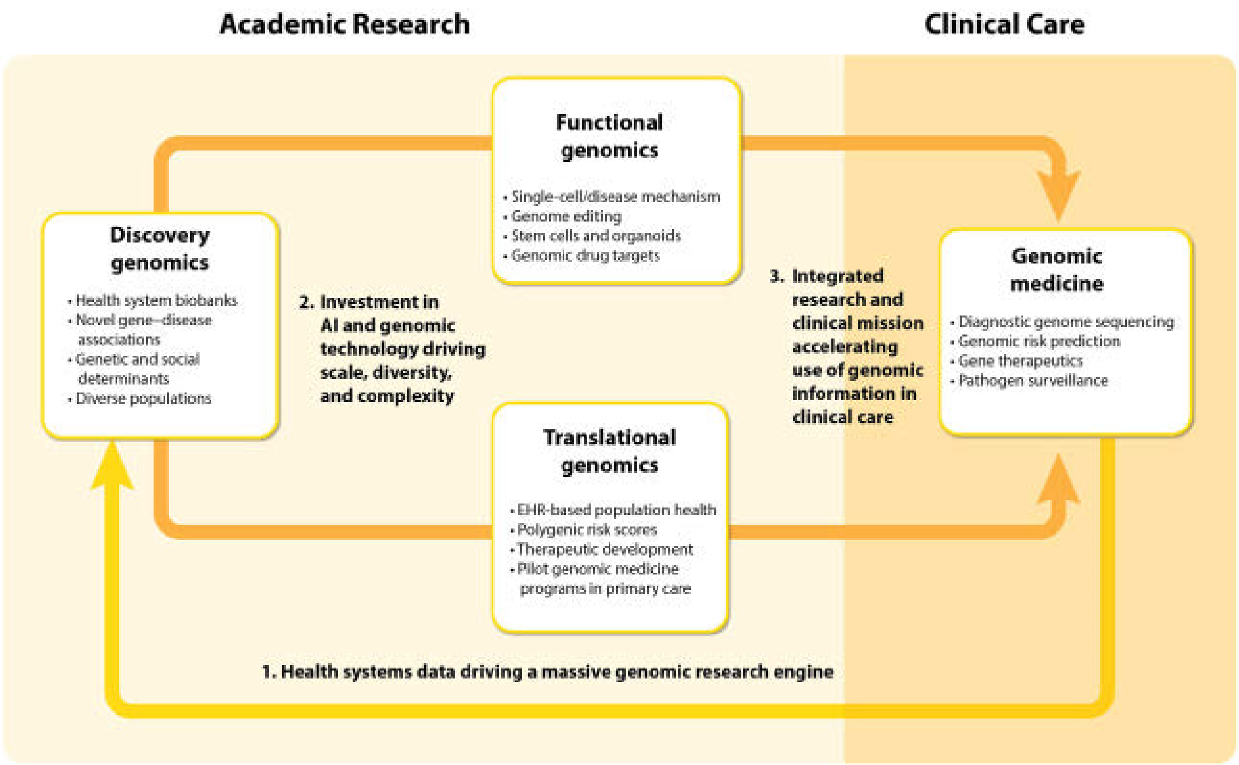

Academic Research
Clinical Care
Functional genomics
• Single-cell/disease mechanism
• Genome editing
• Stem cells and organoids
• Genomic drug targets
Discovery genomics
• Health system biobanks
• Novel gene–disease associations
• Genetic and social determinants
• Diverse populations
2. Investment in AI and genomic technology driving scale, diversity, and complexity
3. Integrated research and clinical mission accelerating use of genomic information in clinical care
Genomic medicine
• Diagnostic genome sequencing
• Genomic risk prediction
• Gene therapeutics
• Pathogen surveillance
Translational genomics
• EHR-based population health
• Polygenic risk scores
• Therapeutic development
• Pilot genomic medicine programs in primary care
1. Health systems data driving a massive genomic research engine

**FIGURE 1. A genomics-informed health system for the future.** EHR, electronic health record.

differences in COVID-19 outcomes (*10*). Large-scale genetic discovery efforts also inform the prioritization of therapeutic targets, providing vital clues to potentially druggable pathways and opportunities for drug repurposing (*11*).

Despite this success, we will face daunting challenges over the coming decade, including issues of scale and diversity. Current global genomic databases are heavily skewed toward representing European ancestry populations (*3*). It is imperative to address this bias to avoid limiting genetic discovery and essential for achieving clinical benefits for diverse patients. To meet this challenge, we and others are further growing discovery genetics into a massive-scale enterprise to ultimately include millions of individuals across the globe. This will serve as a powerful conduit for translating genetic discovery into medical knowledge. For example, we have piloted a program to identify disease-associated genetic variants, which can be used as a screening mechanism for disease prevention and management (*12*). We have also shown that embedded genomic data in health systems can be used to infer genetic ancestry, which can provide evidence-based means to assess—and complement—the use of self-identified race in medicine (*13*). By leveraging our world-class high-performing compute ecosystem, fostering a culture of broad data sharing among our clinical and scientific researchers, and developing sophisticated machine learning approaches (*13*), we will integrate genomic information seamlessly into to our health system in a way that will ameliorate—not promulgate—health disparities.

## Linking genetic discovery to disease mechanisms

Progress in genetic discovery has been paralleled by major advances in functional genomic technology, which systematically interrogates the functional effects of genes and variants on molecular, cellular, and physiological processes. Prominent among these approaches are single-cell analyses, which help elucidate the specific cell populations in which genetic variation operates to influence disease. Single-cell-based transcriptomics have helped elucidate that a genetic driver of Crohn's disease called *NOD2* (first reported 20 years ago) substantially acts via altered blood monocyte differentiation, resulting in impaired myeloid–stromal interactions. We have identified functional gene signatures for treatment resistance to anti-tumor necrosis factor (TNF) therapy (*14*), and for responsiveness to new therapeutic targets [e.g., gp130 (*15*)]. CRISPR/Cas9 screens, in which thousands of genes are perturbed in parallel, offer another major advance. The functionality of CRISPR genomics was recently expanded through the development of a novel protein barcode (Pro-Code) technology, which enables high-dimensional phenotyping of CRISPR screens with single-cell and spatial resolution (*16*). This technology was used to identify two genes, *SOCS1* and *TGFBR2*, which control the immune composition of lung tumors, a major determinant of therapeutic response to immunotherapies (*17*).

It will be important to develop strategies that extend functional genomics to better understand human physiology. Genome editing in induced pluripotent stem cells and differentiation of organoids provide powerful approaches to understand the impact of genetic variation, particularly in tissues that are not easily accessible, such as the brain. For example, we have demonstrated that a variant in the *MAPT* gene is associated with frontotemporal dementia and causes early dysregulation of glutamatergic signaling followed by widespread altered splicing and age-dependent neurodegeneration (*18*). At the tissue level, it will be critical to statistically model biological pathways and cell–cell interactions, including clonal expansion and neoplasia, and cellular recruitment and differentiation. Developing spatial analytic platforms that rigorously and flexibly integrate evolving technical developments is essential.

## Emerging DNA and RNA therapeutics

In synergy with groundbreaking genetics research, scientists are translating insights from the genome into novel gene therapies—medicines that work at the DNA and RNA level—to treat a broad range of diseases, including rare inherited disorders, cancers, and infectious diseases. Decades of research have recently led to gene therapy successes that are life-changing for patients with specific conditions and have the potential to be extrapolated widely to other conditions. For example, Balwani et al. showed in phase 1/2 and 3 clinical trials that givosiran, a short-interfering RNA that works by RNA interference (RNAi), can successfully treat the group of debilitating genetic diseases known as acute hepatic porphyrias, leading to the second FDA-approved RNAi drug (*19*). Genetically engineered T cells have also been established as a highly effective treatment for aggressive blood cancers, and messenger RNA (mRNA) vaccines for COVID-19 have altered the course of the pandemic by providing millions of people with protection from the disease (*20*).

Work is also needed to greatly expand future clinical applications of gene medicine. This includes the development of new lipid nanoparticle technology, which can enhance delivery of mRNA and oligonucleotide drugs, in situ cancer vaccines that use RNA to stimulate the immune system (*21*), and the establishment of new ways to employ CRISPR gene editing for cell engineering. These preclinical and clinical efforts are enabling researchers to generate better and more effective treatments for a wide range of diseases.

## Putting genetic testing at the forefront of clinical care

Important advances are being made in genomic medicine strategies. The NYCKidSeq program is a large clinical trial using genome sequencing as a frontline modality to diagnose disorders that have a complex phenotype, such as neurodevelopmental conditions, in children from diverse and underserved communities in New York City (*22*). We have innovated new digital platforms to facilitate the communication of complex genomic information to diverse and multilingual patients (*23*), and to allow health care providers to engage with clinical genetic laboratories. In parallel, we are rapidly coordinating efforts to integrate more complex genetic risk information (e.g., polygenic risk scores) into preventive health care. As part of the Electronic Medical Records and Genomics (eMERGE) Network, we are offering a Genome-Informed Risk Assessment, including clinical polygenic risk scores for 10 common conditions, to patients in primary care practices with recommendations for how to reduce their risk. Long-term outcomes from these programs will broadly inform the utility and cost-effectiveness of utilizing genomic sequencing as a frontline test for preventive health. Pathogen surveillance is also an important area of genomic medicine, particularly to identify new threats and the emergence of potential deadly variants, such as for flu and SARS-CoV-2 (*24*).

Over time we expect that genome sequencing will increasingly be used to identify disease risk in asymptomatic individuals. Research has shown that genomic risk for medically actionable conditions, including certain cancers and cardiovascular diseases (*25*), is more prevalent than previously estimated, and is largely undiagnosed in routine clinical care. Genomic screening has the potential to identify individuals at highest risk for disease, so that risk-reducing interventions can be offered prior to disease onset. We have implemented a genomic screening program to empower individuals to understand their monogenic disease risks and take actions to reduce them (*12*). Results from the program and from eMERGE are integrated into the electronic health record so that individuals can continue to obtain genome-informed care.

## Conclusions

In the next decade, we expect that clinical genome sequencing will outpace research sequencing, and that genetic information will be centered in routine care throughout the life course of a patient. Through continued investment in genetics, AI, and therapeutics, health care systems will become increasingly informed by genomics (Figure 1), enabling rapid integration of genomic knowledge and therapeutics into clinical practice to benefit patients' lives.

## References

1. E. S. Lander *et al.*, *Nature* **409**, 860–921 (2001).
2. C. M. Eng *et al.*, *N. Engl. J. Med.* **345**, 9–16 (2001).
3. N. S. Abul-Husn, E. E. Kenny, *Cell* **177**, 58–69 (2019).
4. F. K. Satterstrom *et al.*, *Cell* **180**, 568–584.e23 (2020).
5. A. E. Locke *et al.*, *Nature* **518**, 197–206 (2015).
6. J. Flannick *et al.*, *Nature* **570**, 71–76 (2019).
7. J. H. Cho, P. K. Gregersen, *N. Engl. J. Med.* **365**, 1612–1623 (2011).
8. S. M. Purcell *et al.*, *Nature* **506**, 185–190 (2014).
9. I. Landi *et al.*, *Nat. Med.* (2021), doi: 10.1038/s41591-021-01475-7.
10. COVID-19 Host Genetics Initiative, *Nature* (2021), doi: 10.1038/s41586-021-03767-x.
11. B. S. Glicksberg, L. Li, R. Chen, J. Dudley, B. Chen, *Methods Mol. Biol.* **1939**, 91–118 (2019).
12. N. S. Abul-Husn *et al.*, *Genome Med.* **13**, 17 (2021).
13. G. M. Belbin *et al.*, *Cell* **184**, 2068–2083.e11 (2021).
14. J. C. Martin *et al.*, *Cell* **178**, 1493–1508.e20 (2019).
15. S. Nayar *et al.*, *Nature* **593**, 275–281 (2021).
16. A. Wroblewska *et al.*, *Cell* **175**, 1141–1155.e16 (2018).
17. M. Dhainaut *et al.*, *bioRxiv* (2021), doi: 10.1101/2021.07.13.451021.
18. K. R. Bowles *et al.*, *Cell* **184**, 4547–4563.e17 (2021).
19. M. Balwani *et al.*, *N. Engl. J. Med.* **382**, 2289–2301 (2020).
20. S. J. Thomas *et al.*, *N. Engl. J. Med.* (2021), doi: 10.1056/NEJMoa2110345.
21. L. Hammerich *et al.*, *Nat. Med.* **25**, 814–824 (2019).
22. J. A. Odgis *et al.*, *Trials* **22**, 56 (2021).
23. S. A. Suckiel *et al.*, *Genet. Med.* **23**, 942–949 (2021).
24. A. S. Gonzalez-Reiche *et al.*, *Science* **369**, 297–301 (2020).
25. E. R. Soper *et al.*, *J. Pers. Med.* **11** (2021), doi: 10.3390/jpm11010049.

### GERIATRICS AND PALLIATIVE CARE

# Creating new models of care in geriatric and palliative medicine

R. Sean Morrison*, Diane E. Meier, Albert L. Siu

Over the past century, improved living conditions, advances in public health, and modern medical care have nearly doubled human life expectancy and brought with it the gifts of longer life, space to pursue our passions, and more time for deepening relationships with loved ones. These added years—typically after age 75, however—are often accompanied by one or more serious chronic illnesses with which we live for many years before we die. To date, four National Academy of Medicine reports (*1*) and three National Institutes of Health State-of-the-Science workshops/conferences (*2–4*) have reported extensive data from multiple settings and disease categories of the high prevalence of physical, psychosocial, and financial suffering associated with serious chronic illness in older adults. In the coming years, our health care system must pivot to deliver responsive care for older adults living with chronic illnesses and, as recent events have illustrated, address the increased vulnerability of older adults to the effects of inequality, emerging infections, and climate change. The need for evidence-based approaches to make these added years worth having is urgent, as is the need to integrate these approaches into routine practice.

In the United States, health expenditure is projected to be over 30% of the gross domestic product (GDP) by 2040 if growth continues at its current rate. Older adults with serious illness make up only 5%–10% of our population, but account for over two-thirds of all health care spending. In prior work, we have found that older adults living with chronic conditions account for a larger proportion of Medicare costs than those in the last year of life (*5*). Our efforts have focused on development and implementation of new models of care that are matched to the needs of older adults—care that addresses physical and emotional distress, supports family and other caregivers, and maximizes independence and quality of life for as long as possible (*6–11*). We have developed, tested, and widely scaled palliative care—a relatively new medical specialty (*11*). As a result of our efforts, palliative care now reaches >90% of large hospitals in the United States and is spreading rapidly to meet the needs of older and seriously ill people in community settings as well. Research by us and others demonstrates that palliative care improves quality of life, reduces symptom burden, improves family outcomes, reduces need for hospitalization, and, in some populations, prolongs life as compared to traditional care (*12*, *13*). Recognizing that hospitals carry higher risk for older people, the field has also developed, tested, and scaled an important innovation enabling some older adults to receive hospital-level care at home, with support from medical and nursing staff delivered right in the patient's own home (*8–10*). Research demonstrates that this approach provides not only equal or higher quality of care, but also faster recovery and reduced costs.

These new models of care focused on quality of life, maintaining independence, and remaining at home (when possible) are especially important for older adults living with dementia. Alzheimer's disease and related dementias (ADRD) currently affect one in four people over age 80 years and will triple in prevalence in the United States over the next 30 years as people live longer (*14*). Whereas investment in dementia research by the U.S. National Institutes of Health (NIH) and industry has catalyzed our understanding of its pathogenesis, brain structural and functional changes, and behavioral and cognitive manifestations, notable knowledge gaps remain (*15*). Because people with dementia often cannot express their symptoms in words, undertreatment of pain and other symptoms is common. Preventable hospitalizations may occur because of a patient's inability to adhere to treatment and difficulty in recognizing symptoms associated with an impending crisis, magnifying the physical, emotional, and financial demands that dementia places on caregivers. Family caregivers need practical assistance, attention to relief of suffering, and support in order to sustain their caregiving role. Our work assessing service use, costs of care, and patient and caregiver outcomes for the increasing population of persons living with dementia and serious illness will inform new models of care delivery truly matched to the needs of this growing population (*16–20*).

Commonly used statistical methods, such as regression analysis, take a variable-centered approach to data analysis. With these types of statistical approaches, the focus is on learning about relationships between variables. For example, studies have traditionally focused on predictors and outcomes of specific diseases (dementia, heart failure, cancer) or conditions (disability, frailty). The challenge, however, is that these diseases and conditions have qualitatively and quantitatively different patterns of change and probabilities of adverse outcomes over time, as outcomes and their determinants interact with non-health-status factors (*21*). Accurate classification of heterogeneous cases into groups with similar members has tremendous potential for providing information with practical and theoretical use (*22*). Analytic techniques such as cluster analysis and latent class growth analysis provide the opportunity to better understand clinically meaningful patterns of service use across the whole system of care and with the goal of better matching health-services delivery to actual patient need.

Finally, average time from journal publication of new scientific evidence to incorporation into routine clinical practice is 17 years (*23*). Implementation science (IS) is "the scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices (EBP) into routine practice, and, hence, to improve the quality and effectiveness of health services" (*24*). Our experience with national implementation of new models of care for older persons and those with serious illness demonstrates both feasibility and impact. Our focus on the often-underrepresented older persons who need and receive many of these interventions (*25*) points to the necessity of adapting evidence-based interventions to the patients who will receive them. Further, establishing effectiveness is not sufficient. Interventions (e.g., delirium prevention, access to palliative care during serious illness) must be integrated into clinical practice and routinized in order to benefit patients.

---

Brookdale Department of Geriatrics and Palliative Medicine, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: sean.morrison@mssm.edu

Barriers to dissemination and implementation due to behavioral, organizational, payment, or other constraints must be codified and resolved. There is a pressing demographic imperative to attend to dissemination and implementation of models that are matched to the needs of a very different population, now living much longer with serious chronic illnesses, than the one characterizing our country when Medicare and the NIH were first established more than a half-century ago.

## References

1. Institute of Medicine, *Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life* (Institute of Medicine, Washington, DC, 2015).
2. R. S. Morrison, *J. Palliat. Med.* **16**, 726–729 (2013).
3. R. S. Morrison, S. Zieman, *J. Palliat. Med.* **20**, 328 (2017).
4. "NIH State-of-the-Science Conference Statement on Improving End-of-Life Care," *NIH Consens. State Sci. Statements* **21**, 1–26 (2004).
5. M. D. Aldridge, A. S. Kelley, *Am. J. Public Health* **105**, 2411–2415 (2015).
6. A. L. Siu, L. H. Spragens, S. K. Inouye, R. S. Morrison, B. Leff, *Health Aff. (Millwood)* **28**, 113–125 (2009).
7. W. W. Hung, J. S. Ross, J. Farber, A. L. Siu, *JAMA Intern. Med.* **173**, 990–996 (2013).
8. M. R. Augustine *et al.*, *J. Am. Geriatr. Soc.* **68**, 1584–1593 (2020).
9. A. A. Brody *et al.*, *J. Am. Geriatr. Soc.* **67**, 588–595 (2019).
10. A. D. Federman, T. Soones, L. V. DeCherrie, B. Leff, A. L. Siu, *JAMA Intern. Med.* **178**, 1033–1040 (2018).
11. R. S. Morrison, D. E. Meier, *J. Pediatr. Hematol. Oncol.* **33**, S126–S131 (2011).
12. D. Kavalieratos *et al.*, *JAMA* **316**, 2104–2114 (2016).
13. P. May *et al.*, *JAMA Intern. Med.* **178**, 820–829 (2018).
14. Alzheimer's Association, 2021 "Alzheimer's Disease Facts and Figures," *Alzheimers Dement.* **17** (2021).
15. S. L. Mitchell *et al.*, *Ann. Intern. Med.* **156**, 45–51 (2012).
16. K. A. Ornstein *et al.*, *JAMA Intern. Med.* **175**, 1138–1146 (2015).
17. E. J. Pedowitz, K. A. Ornstein, J. Farber, L. V. DeCherrie, *J. Am. Geriatr. Soc.* **62**, 1122–1126 (2014).
18. K. A. Ornstein, K. Boerner, A. L. Siu, R. Schulz, *J. Palliat. Med.* **18**, 736–737 (2015).
19. J. M. Reckrey, L. V. Decherrie, A. S. Kelley, K. Ornstein, *J. Aging Health* **25**, 1036–1049 (2013).
20. C. W. Zhu *et al.*, *Alzheimers Dement.* **11**, 444–454 (2015).
21. L. Lafortune, F. Beland, H. Bergman, J. Ankri, *BMC Geriatr.* **9**, 6 (2009).
22. G. B. Morgan, *Struct. Equ. Model.* **22**, 76–86 (2014).
23. M. S. Bauer, L. Damschroder, H. Hagedorn, J. Smith, A. M. Kilbourne, *BMC Psychol.* **3**, 32 (2015).
24. M. P. Eccles, B. S. Mittman, *Implement. Sci.* **1**, 1 (2006).
25. A. P. Herrera *et al.*, *Am. J. Public Health* **100**, S105–S112 (2010).

---

**HEALTH EQUITY**

# Embedding health equity in biomedical research

**Lynne D. Richardson[1,2,3]*, Gary C. Butts[4,5,6], Carol R. Horowitz[1,3,7]**

The underlying, intersectional causes of health disparities are as varied and complex as health itself, and decades of research uncovering reasons for persistent disparities have not closed the equity gap. The direct impacts of structural racism and bias on Black, Latinx, Native American, and Asian populations have become research and funding priorities, and there is new attention being given to enhancing the diversity of the research workforce. No longer the sole province of a small group of committed researchers, health equity science requires the contributions of scientists from every discipline, working collaboratively with community, clinical, policy, and industry stakeholders (*1*). The challenges and the opportunities are vast.

***Studying intersectional etiologies and multifocal interventions to eliminate disparities.*** Traditional discipline-siloed disparities researchers must dive deeper and search more widely to elucidate and address inequities. Building user-friendly, reliable datasets that link biological, clinical, and social determinants data can improve predictive analytics and enhance understanding of the drivers of health outcomes. Race—a social construct, and ancestry—with its biological components, are both contributors (*2*, *3*). For example, Black/African ancestry adults face a higher burden of kidney failure, driven not only by social determinant and health care factors, but also by a 16-fold increased prevalence of renal-risk *APOL1* genotypes versus White/European ancestry adults (*4*). Social (air pollution) and genetic (*APOL1)* influences significantly interact for overall risk of chronic kidney disease (*5*). Now, a transdisciplinary team is using these findings to influence patient and provider behaviors, develop therapies, and help patients avoid dialysis. Such holistic approaches can realize the promise of big data to investigate how biological, social, financial, neighborhood, and health care factors combine to cause poor health outcomes for racial, ethnic, and gender minorities and socioeconomically and physically disadvantaged groups across a range of diseases and conditions (*6*).

***Dismantling structural racism*** in health care is a prerequisite to achieving equitable health outcomes. This requires rigorous analysis to detect the institutional racism embedded in rules, laws, practices, policies, and organizational approaches. Mount Sinai Health System's new Roadmap for Action uses Collective Impact (*7*) to implement strategies to elevate equity as a core health system value and key quality measure, collect accurate patient demographics, and use informatics to systematically identify and address inequities. For example, inappropriately using race in clinical algorithms based on flawed evidence has caused harm to certain racial groups, leading to well-meaning but misguided calls for a color-blind approach to clinical decision making. However, competently caring for diverse populations requires knowledge of group-specific data on

---

[1]Institute for Health Equity Research, [2]Department of Emergency Medicine, [3]Department of Population Health Science and Policy, [4]Department of Medical Education, [5]Department of Pediatrics, [6]Department of Environmental Medicine and Public Health, [7]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: lynne.richardson@mssm.edu

risks and outcomes and clinical decision tools (*8*). Thus, our system is eliminating use of biased clinical practice and artificial intelligence algorithms through governance structures to use the best available data while articulating and adhering to ethical and equity principles.

***Partnering with communities to improve diversity and inclusion in research.*** Achieving equity will not be possible unless those disproportionately and unjustly impacted by health disparities are partners in conceptualizing, executing, and evaluating research questions, and using findings to inform policy, systems, and environmental changes. Rather than siloing community members in passive advisory groups, they can join other private and public stakeholders as trusted research partners (*9*). Such partnerships can build strategies that overcome the earned skepticism that diverse individuals have of research; inform recruitment and retention strategies so participant demographics in clinical trials reflect the prevalence of conditions being studied; present information in simple, engaging ways; and create positive, respectful research experiences that build a culture of participation (*10*). A common trope is that low-income and nonwhite individuals do not trust researchers (*11*). We challenge researchers to consider another possibility, namely, that they themselves are not sufficiently confident that community members have innovative ideas and strategies to help transform human health.

***Health equity science*** must continue to focus on the pervasive and longstanding disparities in health status and outcomes of diverse populations and investigate how race, ethnicity, language, gender identity, neighborhoods, experiences, and biology lead to illness or wellness (*12*). Achieving health equity will require that scientists across all disciplines learn to bring an equity lens to their work.

## References

1. L. D. Richardson, *J. Public Health Manag. Pract.* **22**, S107–S109 (2016).
2. C. R. Horowitz, *Genet. Med.* **22**, 1461–1463 (2020).
3. C. R. Horowitz *et al.*, *Am. J. Hum. Genet.* **104**, 1088–1096 (2019).
4. A. Parsa *et al.*, *N. Engl. J. Med.* **369**, 2183–2196 (2013).
5. I. Paranjpe *et al.*, *Clin. J. Am. Soc. Nephrol.* **15**, 401–403 (2020).
6. B. Kaplan *et al.*, *Prog. Community Health Partnersh.* **11**, 161–165 (2020).
7. J. Kania, M. Kramer, *Stanford Soc. Innov. Review*, 36–41 (Winter 2011); available at https://ssir.org/articles/entry/collective_impact.
8. L. D. Richardson, G. Abraham, M. Norris, T. Tavernero, "Multicultural Approach to ED Patients," in *Emergency Department Management*, R. W. Strauss, T. A. Mayer, Eds. (McGraw Hill Education, New York, 2013), pp. 187–192.
9. C. R. Horowitz *et al.*, *J. Environ. Res. Public Health* **14**, 225–230 (2017).
10. C. R. Horowitz *et al.*, *Genet. Med.* **21**, 2364–2370 (2019).
11. M. Smirnoff *et al.*, *AJOB Empir. Bioeth.* **9**, 39–47 (2018).
12. L. D. Richardson, M. Norris, *Mt. Sinai J. Med.* **77**, 166–177 (2010).

---

**IMMUNOLOGY**

# Building the future of immunology in medicine

**Miriam Merad[1,2,*], Jeremiah Faith[1,3], Dusan Bogunovic[1,2], Nina Bhardwaj[1,2], Diego Chowell[1,2], Nicolas Vabret[1,2], Brian D. Brown[1,3*]**

## Immunity: In sickness and in health

The immune system maintains human health and plays a central role in virtually all diseases. Protective immune defenses against cancers and infections are critical for survival. However, immune system overactivity can drive detrimental inflammation that causes allergy as well as autoimmune and autoinflammatory diseases such as multiple sclerosis (MS), rheumatoid arthritis (RA), and inflammatory bowel disease (IBD) (*1*). Inflammation is also a major contributor to other conditions, such as atherosclerosis, stroke, Alzheimer's disease, and aging.

Drugs that target the immune system are revolutionizing medicine. Immunotherapy using inhibitors of receptors that stop T cells from killing cancer cells, called checkpoint blockade, and chimeric antigen receptor (CAR) T cells engineered to recognize tumor-associated antigens, have transformed oncology by putting patients with aggressive tumors in long-term remission (*2*). Antibodies that block inflammatory tumor necrosis factor (TNF) effectively treat RA and IBD (*1*), and drugs that deplete B cells reduce inflammatory brain lesions and clinical relapses in MS patients (*3*).

However, a major challenge of "drugging" the immune system is carefully balancing immune activation versus suppression. This conundrum is evident in COVID-19 treatment, as steroids that dampen hyperinflammation are lifesaving, but may prevent virus clearance if used too early in infection (*4*). Thus, we urgently need to identify the mechanisms that drive, brake, and balance immune responses, and understand how to exploit these mechanisms to precisely control immunity and inflammation in diseases.

Technological breakthroughs and new paradigms of multiscale science are addressing this problem. Medical centers have established cross-disciplinary institutes, such as Mount Sinai's Precision Immunology Institute (PrIISM), to study patient tissues, create disease models, identify therapeutic targets, develop novel drugs, and conduct innovative clinical trials. Through this multiscale approach, we are progressing toward treating diseases through more precise and individualized therapeutic control of the immune system.

## Generating immune maps of human diseases

In order to design new immunotherapies, we need accurate and comprehensive information about all the immune

---

[1]Precision Immunology Institute (PrIISM), [2]Department of Oncological Sciences, [3]Department of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding authors: miriam.merad@mssm.edu (M.M.); brian.brown@mssm.edu (B.B.)

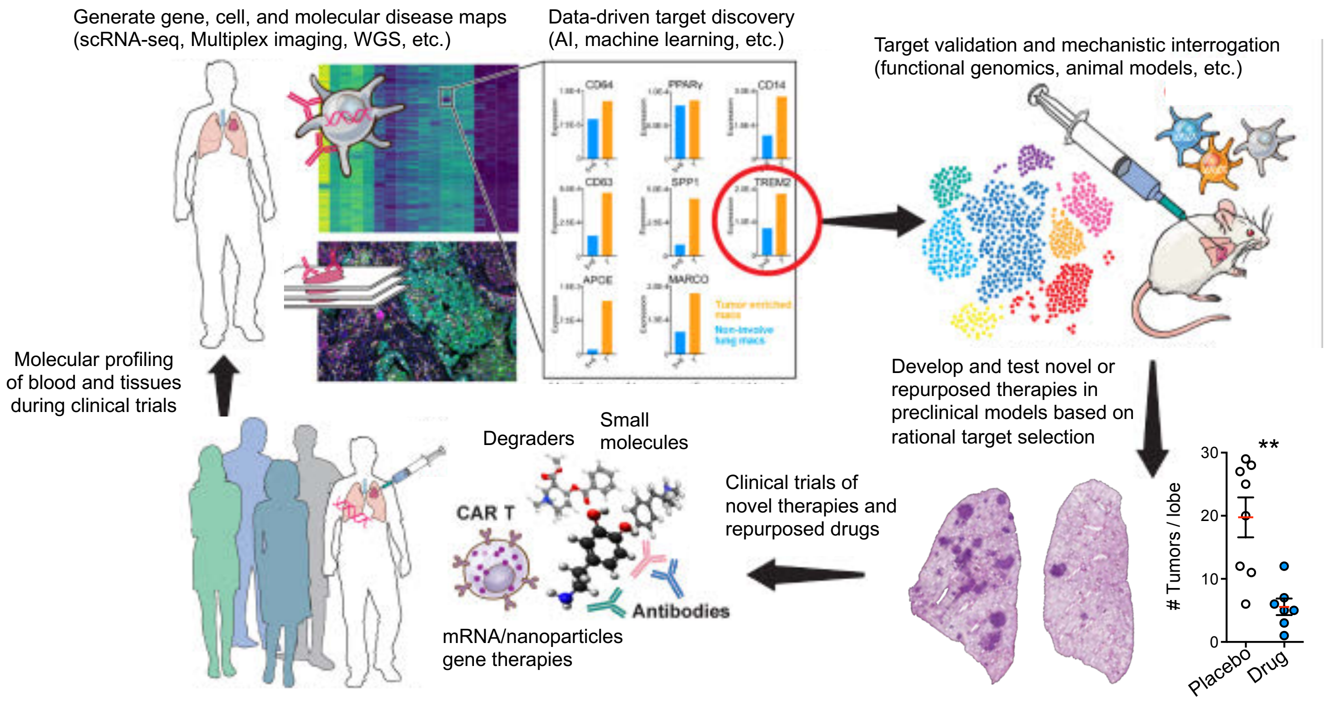

Generate gene, cell, and molecular disease maps
(scRNA-seq, Multiplex imaging, WGS, etc.)
Data-driven target discovery
(AI, machine learning, etc.)
CD64
PPARγ
CD14
CD63
SPP1
TREM2
APOE
MARCO
Expression
Tumor enriched macs
Non-invasive lung macs
Target validation and mechanistic interrogation
(functional genomics, animal models, etc.)
Molecular profiling
of blood and tissues
during clinical trials
Develop and test novel or
repurposed therapies in
preclinical models based on
rational target selection
Degraders
Small
molecules
CAR T
Antibodies
mRNA/nanoparticles
gene therapies
Clinical trials of
novel therapies and
repurposed drugs
# Tumors / lobe
30
20
10
0
**
Placebo
Drug

**FIGURE 1. Building adaptive clinical trial platforms.** scRNA-seq, single-cell RNA sequencing; WGS, whole-genome sequencing.

components of healthy and diseased tissue, how they interact, and the consequences of perturbing these systems via candidate targets. New technological platforms can now measure virtually all molecules, cells, or metabolites within a tissue (*2*). This approach has generated tissue "atlases" with unprecedented detail and revealed specific molecular alterations between healthy and diseased tissue. Immune dysregulation has been mapped at the molecular and cellular level in human lung tumors (*5*), atherosclerosis lesions (*6*), IBD (*7*), COVID-19 (*8*), and allergy (*9*). Among the many findings, these studies have revealed unique molecular programs within macrophages and dendritic cells that dampen tumor immunity and resistance to checkpoint immunotherapy (*10*), identified the potential role of IL-1β in atherosclerotic plaques (*6*), and uncovered a cellular module enriched in IBD patients that resists TNF blockade (*7*). Importantly, these studies revealed actionable targets (*6*, *10*) that enable precise targeting of molecular defects. More work is still needed to enhance these atlases, particularly through incorporation of spatial resolution and longitudinal measurements connecting to health outcomes.

## Immunology meets data science

Data generated by patient analysis can power artificial intelligence (AI) and machine learning algorithms to solve questions about immune responses. For example, a recently developed machine learning model predicts how cancer patients will respond to immune checkpoint inhibitors (*11*). Improved AI models will expedite antigen discovery and vaccine development for seasonal flu, novel coronavirus strains, and other emerging pathogens. Efficient identification of immunogenic antigens among individual tumor mutations is also critical, as personalized cancer vaccines must be administered before the tumor becomes too aggressive. Using computational approaches to sift through the mutational landscape of endometrial, colorectal, and stomach cancers, shared immunogenic frameshift mutations were recently identified between patients (*12*). These findings, and the computational approaches developed in studies like these, will help to establish "off-the-shelf" vaccines for cancer patients.

## Identifying novel immune targets of disease

Through the extensive analysis of patient immune systems, we can identify key variations between healthy and disease states that inform patient treatment. For example, molecular analysis revealed JAK1 (*13*), USP18 (*14*), or TBK-1 (*15*) deficiencies that caused undiagnosed monogenic immune disorders, which led to the repurposing of existing JAK and TNF blockers to precisely treat each patient's condition. Without such a precision approach, these patients would have had no remedy for their debilitating diseases. Functional genomics using CRISPR screens is being employed to great effect to identify novel immunotherapy targets (*16*). Recently, it became possible to resolve CRISPR screens in situ using Pro-Code/CRISPR genomics (*17*). This advancement greatly expands target discovery by enabling the disease-causing role of genes operating extrinsically—such as cytokines—to be determined, and represents a new frontier for identifying immunotherapy targets through functional genomics.

## Developing novel immunotherapies

Immunomodulatory drugs have profoundly improved treatment for many diseases and comprise nine out of the 10 top-selling drugs. However, additional therapies are needed, including for treatment of checkpoint-resistant cancers, atherosclerosis, neuroinflammation, sepsis, and “inflamm-aging.” Many new compounds have already been generated to target the molecules identified from disease profiling and functional genomics studies. A promising example is the macrophage receptor TREM2 (triggering receptor expressed on myeloid cells 2), which was found by molecular analysis in a broad range of diseases, including Alzheimer’s (*18*), IBD (*7*), and lung cancer (*5*); TREM2 blocking antibodies have ameliorated illness in preclinical models of these diseases (*18*) and are currently in clinical trials in cancer patients.

Vaccines were among the first immunotherapies and have had enormous impact on global health crises, including the COVID-19 pandemic. Vaccines educate the immune system to recognize pathogen-associated antigens so it can rapidly block infection by the pathogen. One of the most revolutionary recent developments in this area are messenger RNA (mRNA) vaccines, which are effective not only for viruses but also cancer (*19*). mRNA therapeutics also have applications beyond vaccines. For example, they have been used to reprogram tumor macrophages from protumorigenic to antitumorigenic states (*20*), and to reprogram T cells to kill cancer cells. Scientists are now working to tailor cell type–specific expression of mRNAs (*21*) to enhance targeting of mRNA-based drugs (*22*).

Commensal bacteria have a profound impact on the immune system and human health, including influencing the development of asthma and food allergy (*23*, *24*). Scientists are now determining which microbes control immune function and how, and are using these insights to generate microbiome cocktails to treat IBD and other inflammatory diseases (*24*).

## Adaptive clinical trial design: Molecules and profiling

The same immunomodulatory drug can treat different conditions because the same immune cells reside in all organs and affect different diseases in similar ways. For example, PD1 antibodies are the first U.S. Food and Drug Administration–approved pan-cancer drug, and TNF-alpha inhibitors are used to treat both RA and IBD (*1*). However, physicians have been slow to repurpose drugs indicated for one disease to another. A reason for this is that medical specialties are siloed and study diseases based on the organ of presentation rather than their molecular underpinnings. This extends to how drug trials have traditionally been run and is particularly irrational for diseases driven by the immune system, given commonalities in immune activity between tissues. We and other groups are working to break this old model by developing clinical trials targeting pathogenic molecular programs shared between diseases and by working closely with multidisciplinary teams of immunologists and different medical specialists.

An example of this approach is IL-4 receptor (IL-4R) blockade, which was originally developed to treat asthma. Elevated IL-4 was identified in atopic skin lesions, and clinical trials proved IL-4R blockade was very effective in treating eczema patients (25). Through deep profiling, we recently found a dominant IL-4 signature in the infiltrating immune cells of patient lung tumors and showed that IL-4R blockade could rescue therapeutic immunity in experimental lung cancer that resists checkpoint blockade (*10*). This has prompted a clinical trial to test combined PD-1 and IL-4R blockade in advanced lung cancer (# NCT05013450). These advances outline an adaptive clinical trial design where new knowledge from tissue profiling and therapeutic immunomodulation provide the rationale for the next set of mechanistic studies (Figure 1).

## References

1. I. B. McInnes, E. M. Gravallese, *Nat. Rev. Immunol.* (2021), doi: 10.1038/s41577-021-00603-1.
2. H. Salmon, R. Remark, S. Gnjatic, M. Merad, *Nat. Rev. Cancer* **19**, 215–227 (2019).
3. S. L. Hauser *et al.*, *N. Engl. J. Med.* **358**, 676–688 (2008).
4. The RECOVERY Collaborative Group, *N. Engl. J. Med.* **384**, 693–704 (2021).
5. Y. Lavin *et al.*, *Cell* **169**, 750–765.e17 (2017).
6. D. M. Fernandez *et al.*, *Nat. Med.* (2019), doi: 10.1038/s41591-019-0590-4.
7. J. C. Martin *et al.*, *Cell* **178**, 1493–1508.e20 (2019).
8. D. Mathew *et al.*, *Science* **369**, eabc8511 (2020), doi: 10.1126/science.abc8511.
9. D. Chiang *et al.*, *J. Allergy Clin. Immunol.* **141**, 2107–2120 (2018).
10. B. Maier *et al.*, *Nature* **580**, 257–262 (2020).
11. D. Chowell *et al.*, “An Integrated Clinical-Genetic Model Improves Prediction of Immune Checkpoint Blockade Efficacy across Multiple Cancer Types,” *Nat. Biotechnol.*, in press (2021).
12. V. Roudko *et al.*, *Cell* **183**, 1634–1649.e17 (2020).
13. C. N. Gruber *et al.*, *Immunity* **53**, 672–684.e11 (2020).
14. F. Alsohime *et al.*, *N. Engl. J. Med.* **382**, 256–265 (2020).
15. J. Taft *et al.*, *Cell* **184**, 4447–4463.e20 (2021).
16. R. T. Manguso *et al.*, *Nature* **547**, 413–418 (2017).
17. M. Dhainaut *et al.*, *bioRxiv*, doi: 10.1101/2021.07.13.451021.
18. H. Hampel *et al.*, *Front. Immunol.* **11** (2020), doi: 10.3389/fimmu.2020.00456.
19. U. Sahin *et al.*, *Nature* **585**, 107–112 (2020).
20. F. Zhang *et al.*, *Nat. Commun.* **10**, 3974 (2019), doi: 10.1038/s41467-019-11911-5.
21. B. D. Brown *et al.*, *Nat. Biotechnol.* **25**, 1457–1467 (2007).
22. R. Jain *et al.*, *Nucleic Acid Ther.* **28**, 285–296 (2018).
23. M. Cecilia Berin, *J. Clin. Invest.* **131**, e144994 (2021), doi: 10.1172/JCI144994.
24. G. J. Britton, J. J. Faith, *Annu. Rev. Microbiol.* **75**, 223–242 (2021), doi: 10.1146/annurev-micro-041321-042402.
25. E. Guttman-Yassky *et al.*, *J. Allergy Clin. Immunol.* **143**, 155–172 (2019).

**MATERNAL AND CHILD HEALTH**

# The frontiers of medical research: Child and maternal health

Siobhan M. Dolan[1], Lisa M. Satlin[2,3], Rachel Vreeman[2,4], Bruce D. Gelb[2,3*]

A commitment to ensuring the optimal health and future of every child, adolescent, and pregnant woman worldwide is a priority for scientists in the fields of pediatrics, obstetrics, child development, developmental biology, reproductive health, and global and population health. Breakthroughs in technology, including genome sequencing and artificial intelligence, allow for new scientific discoveries that can be translated into advances in clinical care.

The continuing evolution of genomic medicine will be transformative. Single-gene (Mendelian) disorders, while individually rare, affect ~5% of live births and contribute to roughly 30% of pediatric hospitalizations (*1*). Since the completion of the Human Genome Project and the development of next-generation DNA sequencing (NGS), our ability to sequence individuals' exomes and genomes has been rendered practicable. This, in turn, has accelerated discovery and implementation research in genomic medicine.

With thousands of Mendelian disorders remaining without identified cause, discovery efforts will continue for years (Figure 1). These discoveries will facilitate elaboration of disease pathogenesis, sometimes uncovering unexpected overlap across entities (e.g., histone modifier variants underlying autism and congenital heart disease) (*2*). Gene discoveries will also enable delineation of the relationships of genotypes with phenotypes and clinical outcomes. Sequencing capacity and cost are no longer limiting factors; the focus for the coming years needs to be on leaving no sick child behind. Every family with a child having a potential genetic abnormality must be studied—a model that has worked well in pediatric oncology.

Implementation research is needed in order to fully realize the potential of pediatric genomic medicine. As clinical genetic test menus expand, important questions about clinical utility remain, which, in turn, are limiting the willingness of health care payers to cover these tests. For example, is genetic testing of critically ill infants cost effective and, if so, which test (targeted gene panels versus genomes) has the highest utility? Subsequent research will be needed to determine how best to standardize deployment of NGS across health systems and to support a workforce that can implement genomic medicine to maximize the benefits to affected children.

Attention is also needed to assure that pediatric genomic medicine is deployed equitably. Statistics for large genetic studies and aggregated NGS data such as in the Genome Aggregation Database (gnomAD) reveal an unacceptable bias toward whites of European descent. Through carefully crafted recruitment strategies and collaborative engagement with community partners, enrollment from groups underrepresented in medical research is achievable.

Critical to the future heath of children are efforts to develop, test, and evaluate existing drugs and novel therapies for safe and effective treatment of genetic disorders. For example, U.S. Food and Drug Administration–approved immunomodulatory drugs have been repurposed to treat children with rare Mendelian disorders (*3*, *4*). For most traits, for which pathogenesis is unknown or rational repurposing is not obvious, studies following the traditional bench-to-bedside path, starting with modeling in animals and cells (now including pluripotent stem cell–derived human cells and organoids) and moving to testing of known drugs and novel compounds, will be required. Of note, neurodevelopmental disorders, no longer viewed as largely inevitable defects of early brain development, represent one area of child health that is ripe for therapy development. Finally, indications for genomic therapies—gene therapies and gene editing—will continue to expand impressively (*5*), although many hurdles remain for rendering them practical for the thousands of Mendelian traits.

Precision obstetrical medicine has the potential to transform care during pregnancy. For instance, obstetricians and geneticists at Mount Sinai, in conjunction with data scientists at the health intelligence company Sema4, are developing a point-of-care decision tool to optimize the timing of cell-free DNA aneuploidy screening. Since a minimum threshold of fetal DNA must be present, we are developing predictive models using factors affecting the fetal fraction, such as gestational age, maternal body mass index, and medications that determine each patient's earliest window for screening. These models will allow personalized screening to assure the greatest likelihood of accurate results. Another example of personalized obstetrical care is being developed using digital phenotyping to predict risk and guide care. Scientists are building models using electronic medical record data that will guide intrapartum monitoring based on postpartum hemorrhage risk (*6*), with future efforts aimed at predicting preeclampsia risk.

Child and maternal health priorities demand that we broaden our scope globally. To reach the United Nations Sustainable Development Goals for 2030, we must recognize exactly what the "Every Woman Every Child Global Strategy" for 2030 states: that women, children, and adolescents are "potentially the most powerful agents for improving their own health and achieving prosperous and sustainable societies" (*7*). As highlighted in the COVID-19 pandemic, we must prioritize global strategies and innovations to adequately support maternal and child health needs for all (*8*). Our research strategy in prioritizing these goals is to engage with communities around socially acceptable, culturally adapted, and validated approaches for detecting growth or developmental challenges and implementing referral strategies to address issues early (*9*, *10*). We must also continue research to detect and address children's long-term health needs, supporting children and their families through lifelong treatment regimens (*11*, *12*), for example, living with HIV, and addressing the broad range of behavioral and social health needs demonstrated by adolescents coping with health challenges globally, such as the cohort of adolescents Mount Sinai has followed for more than 10 years in Kenya (*13*–*16*).

While the COVID-19 pandemic has been challenging clinically, it has also provided research opportunities in child and maternal

---

[1]Department of Obstetrics, Gynecology, and Reproductive Science, [2]Department of Pediatrics, [3]Mindich Child Health and Development Institute, [4]Arnhold Institute for Global Health, Department of Global Health and Health Systems Design, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: bruce.gelb@mssm.edu

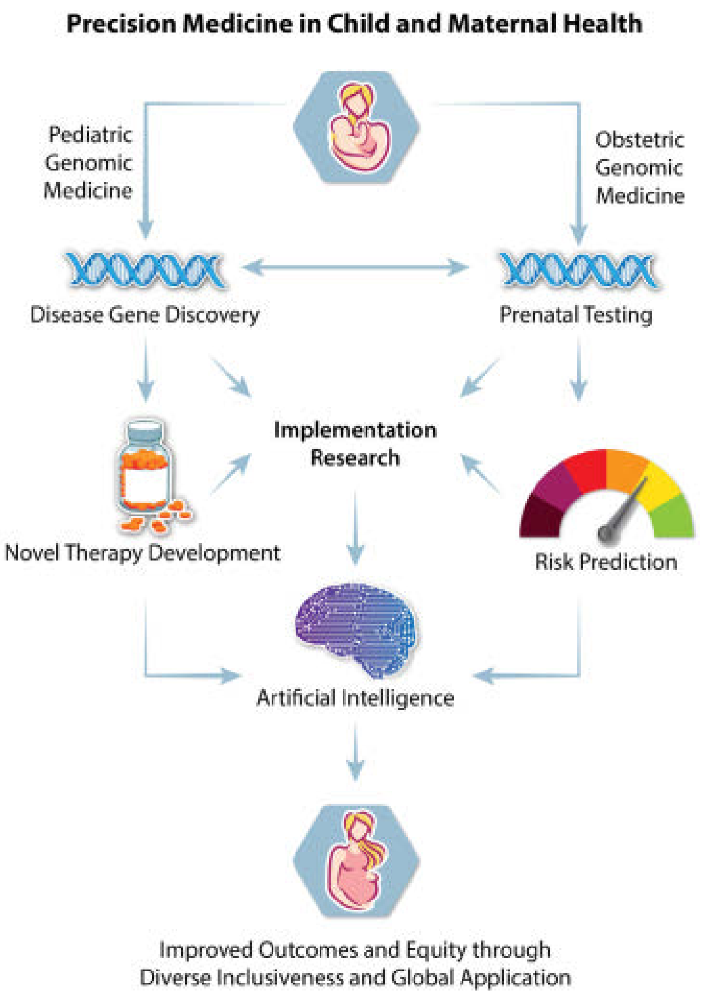

Precision Medicine in Child and Maternal Health
Pediatric Genomic Medicine
Obstetric Genomic Medicine
Disease Gene Discovery
Prenatal Testing
Implementation Research
Novel Therapy Development
Risk Prediction
Artificial Intelligence
Improved Outcomes and Equity through Diverse Inclusiveness and Global Application

**FIGURE 1. Schematic summarizing the future of research in child and maternal health.**

health. Using lessons learned in low-resource settings abroad, we studied universal screening for SARS-CoV-2 infection among pregnant women in Queens, New York (*17*) and assessed childbirth satisfaction and postpartum health during the pandemic (*18*). We also launched a pregnancy cohort to study the short- and long-term impact of COVID-19 on pregnancy (*19*). The Generation C (Generation COVID) cohort captures questionnaire data and biospecimens throughout pregnancy and the postpartum period. Maternal immune responses are being assessed via cytokine levels in blood and cerebrospinal fluid and correlated with pregnancy outcomes, maternal mental health, and child development. New assays are being developed to distinguish the immunity provided by previous infection versus immunization, and newborn blood spots are being correlated to assess immune status at birth. Additional analyses are examining how structural racism and pandemic stress are associated with SARS-CoV-2 infection, birth outcomes, and vaccine deliberation among pregnant women. In the future, we will study the long-term impact of the pandemic on the mental health of mothers and children.

In summary, we expect the frontiers of maternal and child health research to range from improving precision/genomic medicine, to driving disease discovery and generating novel therapeutics, to analyzing population-level data globally. The resulting advances promise to improve health outcomes for women, children, and their families. With proper attention, we can and must assure that these benefits are enjoyed equitably in the United States and throughout the world.

## References

1. C. R. Ferreira, *Am. J. Med. Genet. A* **179**, 885–892 (2019)..
2. J. Homsy *et al.*, *Science* **350**, 1262–1266 (2015).
3. C. N. Gruber *et al.*, *Immunity* **53**, 672–684.e11 (2020).
4. J. Taft *et al.*, *Cell* **184**, 4447–4463.e20 (2021).
5. E. E. Kenny, N. S. Abul-Husn, B. D. Brown, J. H. Cho, A. M. Goate, "Genetics and Gene Therapeutics: Igniting a Revolution in Health Care," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 16–18.
6. A. B. Zheutlin *et al.*, *J. Am. Med. Inform. Assoc.*, ocab181 (2021), doi: 10.1093/jamia/ocab181.
7. Every Woman Every Child, *The Global Strategy for Women's, Children's and Adolescents' Health (2016-2030)* (2015), p. 50; available at EWEC_Global_Strategy_EN_inside_LogoOK_web.pdf (everywomaneverychild.org).
8. M. Ballard *et al.*, *BMJ Glob. Health* **5**, e002550 (2020).
9. E. Oyungu, A. Roose, A. R. Ombitsa, R. C. Vreeman, M. S. McHenry, *Int. J. MCH AIDS* **10**, 128–133 (2021).
10. M. S. McHenry *et al.*, *Res. Dev. Disabil.* **110**, 103837 (2021).
11. R. C. Vreeman, B. M. McCoy, S. Lee, *J. Int. AIDS Soc.* **20**, 21497 (2017).
12. R. C. Vreeman *et al.*, *PLOS One* **13**, e0191848 (2018).
13. L. A. Enane *et al.*, *AIDS Care* **32**, 104–112 (2020).
14. L. A. Enane *et al.*, *J. Int. AIDS Soc.* **24**, e25695 (2021).
15. A. Chory *et al.*, *AIDS Behav.*, 1–11 (2021), doi: 10.1007/s10461-021-03376-9.
16. Collaborative Initiative for Paediatric HIV Education and Research (CIPHER) Global Cohort Collaboration, *J. Int. AIDS Soc.* **23**, e25477 (2020).
17. S. Maru *et al.*, *PLOS One* **15**, e0238409 (2020).
18. T. Janevic *et al.*, *Matern. Child Health J.* **25**, 860-869 (2021).
19. N. Molenaar *et al.*, *medRxiv* (2021), doi: 10.1101/2021.02.01.21250943.

**MICROBIOLOGY**

# Preparing for the next pandemic

**Florian Krammer[1,2,*], Viviana Simon[1,2,3], Adolfo García-Sastre[1,2,3,4,5], Peter Palese[1,4]**

Early in January 2020, the sequence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was published online by Chinese scientists from Fudan University in Shanghai. Scientists in the Department of Microbiology and in the Global Health and Emerging Pathogens Institute at the Icahn School of Medicine at Mount Sinai were alarmed and started to prepare even before it was clear that the cluster of patients with respiratory disease first reported in Wuhan in 2019 would be the start of a pandemic triggering an unprecedented global health crisis. Within a few weeks, molecular testing for the virus was established; an antibody test was developed suitable for studying the immune response after SARS-CoV-2 infection and identifying possible COVID-19 (coronavirus disease 2019) survivors who might be eligible for plasma donations (*1–4*); and animal models (*5*, *6*) and antiviral and vaccine research for COVID-19 were initiated (*7*, *8*).

In the following months of 2020, our groups led numerous studies aimed at understanding the biology and pathogenicity of the new virus, and established observational longitudinal cohorts to better understand immunity induced after infection (*9*, *10*). We shared reagents generated in our laboratories at Mount Sinai with hundreds of laboratories and hospitals worldwide, and we spearheaded the development of affordable COVID-19 vaccines (*11*). The reason we could rapidly spring into action and generate these data so quickly was because we were prepared: We had the expertise and necessary infrastructures in place as well as unwavering support from our institution.

Traditionally, our Department of Microbiology focuses on studying RNA viruses and the development of vaccines and therapeutics. We lead one of the Centers of Excellence for Influenza Research and Response (CEIRR), which focuses on pandemic preparedness. The center contributed significantly to responding to the 2009 H1N1 pandemic and several outbreaks of zoonotic influenza viruses, including H7N9 (Figure 1). We also lead a Collaborative Influenza Vaccine Innovation Center (CIVIC) that is focused on developing universal influenza virus vaccines, discussed further below. This past experience with influenza virus and other RNA viruses, along with close collaborations with other research and clinical groups on campus, allowed us to respond rapidly to SARS-CoV-2. For example, our Personalized Virology Initiative and Pathogen Surveillance Program are both longstanding multidisciplinary programs that allowed immediate access to critical patient samples and virus isolates collected in real time with approval of the Institutional Review Board.

We are now 20 months into the COVID-19 pandemic. While new infections have decreased and vaccines have become widely available in high-income countries, much remains to be learned and accomplished in order to crush COVID-19 and prevent future pandemics. Current high-profile projects aim to provide a better understanding of vaccine- and infection-induced immunity (*12*, *13*), to define correlates of protection for SARS-CoV-2, and to explain why some viral variants (e.g., the B.1.617.2 Delta variant) are more transmissible and can cause infections in fully immunized individuals (*14*). To answer these questions, our group leads the Protection Associated with Rapid Immunity to SARS-CoV-2 (PARIS) program, which follows infected and vaccinated individuals over time to study reinfection rates and the longevity of immune responses. This has been complemented with the rapid establishment of in vitro and in vivo models of SARS-CoV-2 infection under BSL3 containment. Multi-institutional collaborations were also established, allowing the rapid discovery of potential antivirals and treatments, several of which, including plitidepsin and topotecan, have been advanced into clinical trials (*15*, *16*). Finally,

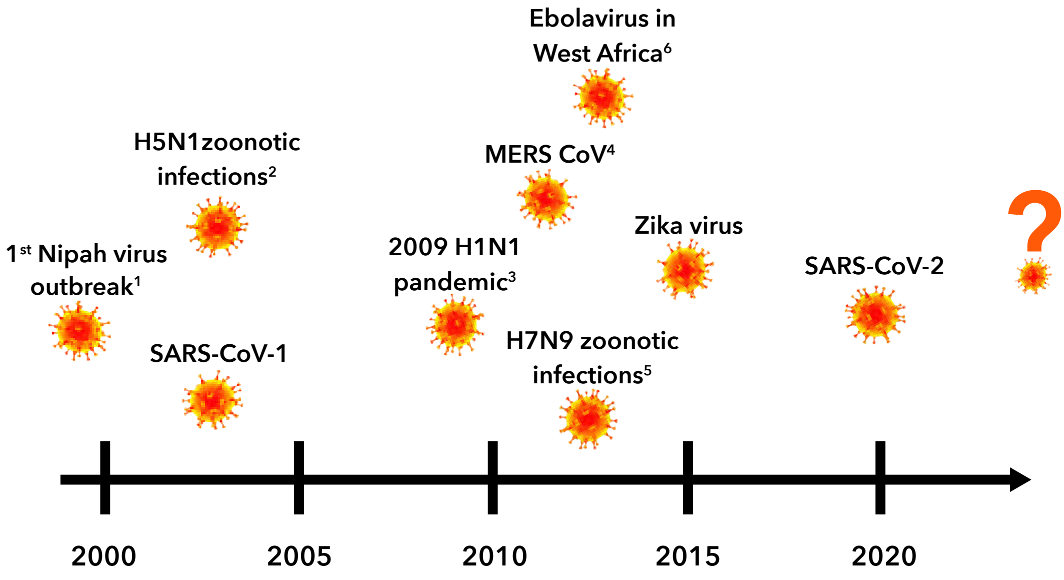

Ebolavirus in West Africa[6]
H5N1zoonotic infections[2]
MERS CoV[4]
Zika virus
1st Nipah virus outbreak[1]
2009 H1N1 pandemic[3]
SARS-CoV-2
?
SARS-CoV-1
H7N9 zoonotic infections[5]
2000
2005
2010
2015
2020

**FIGURE 1. Major outbreaks of emerging viruses in the last 20 years.** [1]Nipah virus caused its first detected outbreak in 1998/1999, but since then has caused several additional outbreaks. [2]The highly pathogenic H5N1 virus first caused an outbreak in 1997, then reemerged in 2003 and has caused sporadic zoonotic infections since then. [3]Pandemic H1N1 emerged in 2009 and has since become a seasonal influenza virus. [4]MERS CoV was first detected in 2012 and has caused sporadic outbreaks since then. [5]H7N9 virus first caused an outbreak in 2013, and since then has caused sporadic outbreaks. [6]Ebolavirus caused a large outbreak in West Africa from 2013 to 2016, but caused many sporadic outbreaks before and after that epidemic. Avian and swine influenza viruses of different subtypes constantly cause zoonotic infections that could lead to larger outbreaks and pandemics.

---

[1]Department of Microbiology, [2]Department of Pathology, Molecular and Cell-Based Medicine, [3]Global Health and Emerging Pathogens Institute, [4]Department of Medicine, Division of Infectious Diseases, [5]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: florian.krammer@mssm.edu

we are spearheading the development of a Newcastle disease virus (NDV)-based SARS-CoV-2 vaccine that can be produced using existing influenza vaccine production plants at very low cost (*17*). This vaccine is currently in clinical development in Thailand, Vietnam, Brazil, and Mexico (*18*). We hope that this work will make SARS-CoV-2 vaccines available in low- and middle-income countries, especially since it can be produced at low cost locally—providing countries independence from international vaccine supplies.

The efforts mentioned above focus on the current pandemic, but it is also clear that we need to plan ahead. While we do not know which virus will cause the next pandemic or when it will occur, it is certain that new viral pathogens will continue to spill over into human populations, giving rise to future pandemics, unless safeguards are put in place now. Increase in farming, habitat destruction, extreme weather conditions, and global warming will likely result in an increased frequency of pandemics. Four of the last five pandemics with respiratory viruses have been caused by influenza virus, and the high diversity of this virus in the animal reservoir makes it a very likely candidate for the next pandemic. To tackle this problem—and to solve the issue of antigenic drift of seasonal influenza viruses—we are developing several broadly protective and universal influenza virus vaccine candidates, of which several are in clinical development (*19*). We have also developed the NDV vaccine platform stated above, which can be used to generate a wide variety of vaccines, including those against respiratory syncytial virus, filoviruses, influenza virus, and any emerging viral pathogen.

Furthermore, we recently established the Center for Vaccine Research and Pandemic Preparedness (C-VARPP). C-VARPP focuses on collecting biospecimens from patients seeking care at the Mount Sinai Health System, the development of monoclonal antibody therapeutics from convalescent individuals, and the analysis of infection- and vaccine-induced immunity. New York City (NYC) is a major travel hub, and individuals from many different regions of the globe live here or come to visit. Moreover, people in NYC, like those in other large metropolitan areas, live in close proximity and rely on public transportation to commute. This constant influx of people infected with emerging, reemerging, and circulating pathogens such as Zika virus, dengue virus, SARS-CoV-2, and others, not only compels our health system to provide innovative, high-quality diagnostics and clinical care, but also allows us to seamlessly integrate cutting-edge research aimed at studying and fighting these pathogens (*20–25*). Innovative programs such as these require robust funding from governmental and private sources.

Many challenges remain with endemic viruses such as seasonal influenza virus, and most importantly, the global community recently had to discover that we are not prepared for emerging pandemic viruses. We must focus on studying viruses and developing suitable countermeasures such as vaccines and therapeutics. The scientific enterprise was prepared to take on COVID-19, which enabled a rapid understanding of the virus and development of treatments and remarkably effective vaccines. While the pandemic is not over yet and still needs our full attention, work should continue on solving remaining problems with endemic and pandemic viruses in order to be better prepared for the next outbreak—which will certainly come.

## References

1. F. Amanat *et al.*, *Nat. Med.* **26**, 1033–1036 (2020).
2. D. Stadlbauer *et al.*, *Curr. Protoc. Microbiol.* **57**, e100 (2020).
3. F. Krammer, V. Simon, *Science* **368,** 1060–1061 (2020).
4. J. M. Carreño *et al.*, *iScience* **24**, 102937 (2021).
5. R. Rathnasinghe *et al.*, *Emerg. Microbes Infect.* **9**, 2433–2445 (2020).
6. F. Amanat *et al.*, *mBio* **12** (2021), doi: 10.1128/mBio.02648-20.
7. W. Sun *et al.*, *Vaccines (Basel)* **8**, 771 (2020).
8. W. Sun *et al.*, *EBioMedicine* **62**, 103132 (2020).
9. A. Wajnberg *et al.*, *Science* **370**, 1227–1230 (2020).
10. D. Stadlbauer *et al.*, *Nature* **590**, 146–150 (2021).
11. F. Krammer, *Nature* **586**, 516–527 (2020).
12. F. Krammer *et al.*, *N. Engl. J. Med.* **384**, 1372–1374 (2021).
13. O. Van Oekelen *et al.*, *Cancer Cell* **39**, 1028–1030 (2021).
14. J. M. Carreño *et al.*, "Evidence for Retained Spike-Binding and Neutralizing Activity Against Emerging SARS-CoV-2 Variants in Serum of COVID-19 mRNA Vaccine Recipients," *EBIOMedicine*, in press (2021).
15. K. M. White *et al.*, *Science* **371**, 926–931 (2021).
16. J. S. Y. Ho *et al.*, *Cell* **184**, 2618–2632.e17 (2021).
17. W. Sun *et al.*, *bioRxiv* (2021), doi: 10.1101/2021.07.06.451301.
18. ClinicalTrials.gov, IDs NCT04764422, NCT04830800, NCT04993209 and NCT04871737.
19. R. Nachbagauer *et al.*, *Nat. Med.* **27**,106–114 (2020).
20. F. Amanat *et al.*, *Cell* **184**, 3936–3948.e10 (2021).
21. M. J. Bailey *et al.*, *Nat. Commun.* **9**, 4560 (2018).
22. J. A. Brown *et al.*, *Immunity* **50**, 751–762.e5 (2019).
23. S. V. Bardina *et al.*, *Science* **356**, 175–180 (2017).
24. A. S. Gonzalez-Reiche *et al.*, *Science* **369**, 297–301 (2020).
25. M. M. Hernandez *et al.*, *Nat. Commun.* **12**, 3463 (2021).

---

**NEUROSCIENCE**

# The human brain: Medicine's last frontier

**Eric J. Nestler[1,2]*, Yasmin L. Hurd[1,2], Paul J. Kenny[1], Rene Kahn[2], Barbara Vickrey[3]**

The human brain is uniquely complex. A single brain is composed of ~100 billion neurons and roughly an equal number of glia, with perhaps ~100 trillion synapses comprising the brain's neural circuitry. While computers are faster and more accurate, a human brain still outperforms the most advanced computer in abstract thinking, planning, social intuition, and emotion. As progress is made in treating and preventing diseases of other organs, and people live longer, the impact of brain disorders continues to increase, today representing five of the top 10 causes of disability worldwide (depression, psychosis, dementia, stroke, and addiction). Unlike all other branches of medicine, where much of today's challenges will be met by executing existing approaches, the brain still requires fundamentally new paradigms to understand its functioning in health and disease. We are entering a phase of unprecedented optimism and anticipation based on a range of exciting new tools that have been developed over the past decade. This is reflected in major investments in brain science across the globe, such as the U.S. BRAIN (Brain Research Through Advancing Innovative Neurotechnologies) Initiative and related efforts in Europe and Asia.

The brain's extraordinary complexity resides in its cellular diversity and highly intricate circuits. Each brain cell has the same complexity of molecular constituents as all cells in peripheral tissues. However, in contrast to those tissues, whose cellular function in vitro reflects the organ's function in vivo (e.g., pumping blood, filtering air or blood, attacking pathogens, or secreting insulin), the brain exhibits so-called "emergent properties" that cannot be understood based on molecular–cellular function alone. While much attention is focused on consciousness and "the mind"—which can be viewed as the ultimate emergent properties—even the most basic functions of the brain, such as interpreting a visual image, throwing a baseball, solving a mathematical equation, remembering past experiences, or emoting in response to good or bad news, cannot be explained today by simply measuring cell and circuit activity.

## Technology-driven roadmap for understanding the brain

These challenges define an overall experimental strategy that is now being used to decipher brain function under normal conditions and resolve what goes wrong in specific disease states (Figure 1). Starting with molecular biology—analogous to other fields and relying on their tools—it is now possible to understand how a cell's molecular constituents are determined initially by an individual's genetic constitution and modified across the lifespan via epigenetic processes during development and in response to lived experiences and other environmental factors throughout life. Such molecular characterization must be achieved with anatomical precision, by use of single-cell RNA-sequencing (scRNAseq) and related methods, followed by "spatial transcriptomics," which permits delineation of the molecular constituents of a cell within the context of its location in the brain and its circuitry (*1*). New technologies are required to better link scRNAseq with spatial information such that the transcriptomes of individual neurons can be related directly to their unique patterns of connectivity in the brain. Other new methods such as iDISCO (immunolabeling-enabled three-dimensional imaging of solvent-cleared organs) or CLARITY (Clear Lipid-exchanged Acrylamide-hybridized Rigid Imaging/Immunostaining/in situ-hybridization-compatible Tissue hYdrogel) enable mapping proteins in a brain-wide fashion with cellular resolution (*2*). These tools are being applied not only to rodent models but also to human brain tissue obtained at autopsy or during surgery (*3*). scRNAseq is revealing an unexpected level of cellular diversity in the brain by identifying numerous putative subtypes of a previously defined single cell type, with work now underway to understand which such subtypes reflect distinct functional states of a single cell type versus truly distinct types of cells. This work is also generating newfound appreciation of the importance of glia and endothelial cells in neuropsychiatric phenomena. A related challenge is to understand the specification of such a large number (many thousands) of distinct neuronal and nonneuronal cell types through development.

These molecular constituents determine the functioning of individual cells, which in turn control the functioning of simple and more complex composite circuits in the brain and spinal cord that produce behavioral outputs, whether interpreting sensory information; executing a motor program; or learning, remembering, planning, and emoting. The last decade has seen revolutionary advances in our ability to connect the functioning of a single circuit—with a precision and specificity unimagined even two decades ago—to a specific behavior. For example, in vivo $Ca^{2+}$ imaging provides continuous activity measures of thousands of individual neurons in behaving animals (*4*), while optogenetics and chemogenetics make it possible to link the activity of clusters of precisely defined neurons and their projections to a behavior (*5*, *6*). Circuit control over behavior is more complicated than anticipated, in part because each "microcircuit" (one cell type projecting to a second cell type) is modified by both cell types' other projections, which shape the ultimate behavioral output (*7*). Advances in neural network analysis are beginning to reveal how collections of neurons can generate a simple behavior (*8*), although this area requires further innovations through artificial intelligence and machine learning approaches. Importantly, every time a circuit functions, the type and intensity of activity alters the molecular constituents of individual cells in the circuit, which in turn alters the functioning of those circuits. Through this constant iterative plasticity across cells, synapses, circuits, and behavioral responses and back, the brain is ever changing and adapting. Establishing causal connections bidirectionally across this continuum is for the first time technically possible and represents one of the major goals of contemporary neuroscience.

## Progress in delineating the pathophysiology of brain diseases

As we learn more about the brain, it is now possible to better elucidate the causes and pathophysiology of brain and nervous system disorders, although progress is clearly lagging behind that for other medical disciplines due to the enormous complexity involved. Most brain disorders are highly heritable, although with genetic underpinnings far more complex than ever imagined: Many hundreds of genome variations, each contributing a miniscule fraction, determine an individual's disease risk. This situation is similar to the genetic basis of other, common complex disorders. A major challenge is how to use this genetic information to improve diagnosis, for example, through polygenic risk scores. Genetic knowledge is also providing new biological insight into disease pathogenesis, for example in psychiatry, by demonstrating a ~50% overlap of genes implicated in schizophrenia also being seen in either bipolar disorder or autism, as compared to eating disorders (e.g., anorexia nervosa), where genes related to peripheral metabolism are especially implicated (*9*). The genetic complexity of brain disorders emphasizes the potential utility of studying

---

Departments of Neuroscience[1], Psychiatry[2], and Neurology[3]
Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: eric.nestler@mssm.edu

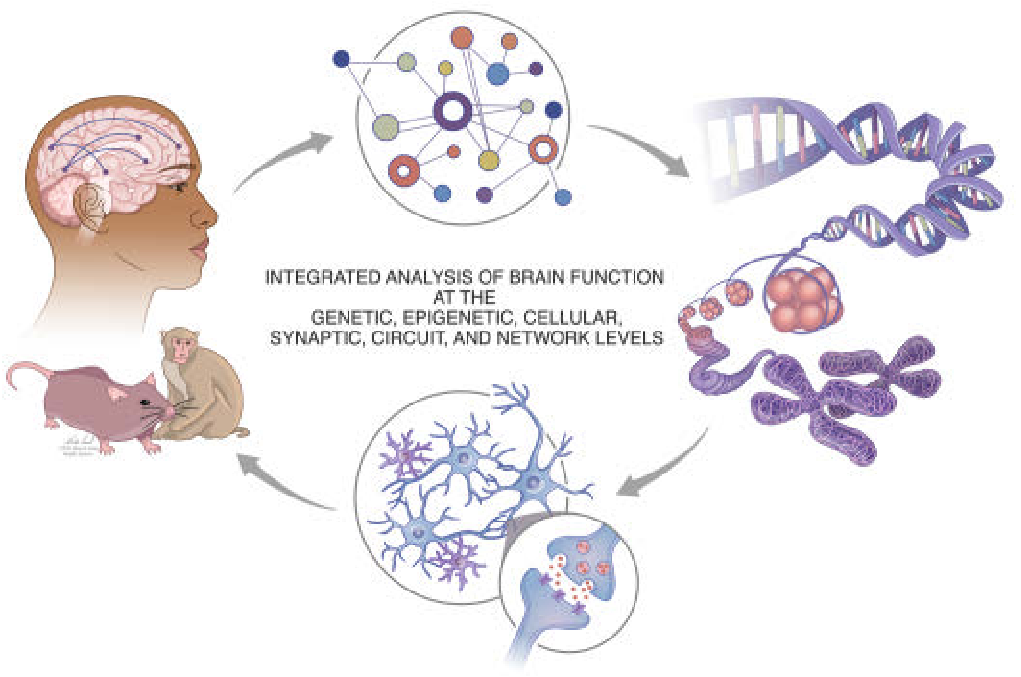

INTEGRATED ANALYSIS OF BRAIN FUNCTION
AT THE
GENETIC, EPIGENETIC, CELLULAR,
SYNAPTIC, CIRCUIT, AND NETWORK LEVELS

**FIGURE 1. A roadmap for understanding brain function in health and disease.** Genetic and epigenetic mechanisms (**right panel**) determine the functioning of individual neuronal (blue) and nonneuronal (purple) cells and synapses (inset) (**lower panel**), which control the functioning of those neurons in larger brain circuits across species (**left panel**). Complex neural networks (**upper panel**) ultimately underlie behavioral responses as well as feedback and alter the molecular and cellular properties of individual cells. Image from Ni-ka Ford (Icahn School of Medicine at Mount Sinai).

neuron- or glia-like cells induced from human patients, which provide the full complement of genetic risk in a cell model. Genomic and transcriptomic data are likewise demonstrating the validity of animal models: Rather than postulating how a rodent behavior may be similar to a human behavior, it is now possible to determine the extent to which an animal model recapitulates the molecular pathology of a human disorder in a cell-type-specific manner. The importance of epigenetic mechanisms in disease is underscored by the large number of genes encoding chromatin regulatory proteins that have been linked to psychiatric and neurological disorders (e.g., *10–12*). Medications targeting epigenetic mechanisms—an intense area of interest in cancer—are as a result now being explored for a range of brain disorders as well. Furthermore, while the clinical application of gene editing and epigenome editing approaches to brain disorders (*13*) will come long after their application to non-nervous system diseases, due to challenges of accessing the brain and limiting side effects, important strides are being made, with nucleic acid–based therapeutics for neurodegenerative disorders showing great promise in the clinic (*14*).

Circuit-based therapeutics represent another enormous opportunity. Deep brain stimulation (DBS) is now widely used to treat Parkinson's disease and other movement disorders, and is being applied to several psychiatric syndromes, including obsessive-compulsive disorder (OCD) and Tourette syndrome, and experimentally to depression, addiction, pain, Alzheimer's disease, and eating disorders (*15*, *16*). DBS involves neurosurgery and therefore is only applicable today for individuals with severe disease, but the surgery is extremely well tolerated, and one can imagine broader applicability as the specificity of treatment is improved. For example, will it be possible to stimulate only an individual cell type in a targeted brain region by employing an optogenetic or chemogenetic approach? Beyond therapeutics, the human neurosurgical suite is an essential arena for researching how circuit function generates behavior, including highly complex outputs such as language, emotion, and memory (*16*, *17*). This work highlights the importance of research on nonhuman primates. While increasingly complex behavioral paradigms are being developed for rodents, there is a limit to how much can be learned about the human brain from rodents, given the extraordinary evolution of frontal cortical circuits in primates (*18*). In fact, the interplay between humans and nonhuman primates offers a critical opportunity (Figure 2). Continuous recordings from DBS electrodes, and from indwelling electrodes in the temporal lobe of epilepsy patients, are integrated with recordings from homologous monkey circuits, with hypotheses about causality between a circuit and a behavioral output tested in the latter. Molecular genetic control of circuit function in primates represents a tremendous challenge, but is perhaps feasible through the study of smaller primates with rapid breeding cycles, such as marmosets. Advanced brain imaging, which offers far broader applicability, is also being used to study brain circuitry in humans and nonhuman primates: While imaging methods still lack the molecular and cellular resolution possible with more invasive tools, improvements in spatial and molecular resolution are continuously being achieved.

## Brain–body interactions

As more is learned about the brain, it is increasingly possible to delineate bidirectional interactions between the brain and peripheral organs, and this synergy is captured perfectly by the multidisciplinary Friedman Brain Institute at Mount Sinai. There is tremendous excitement in neuroimmunology. For example, the last decade has seen major advances in treating autoimmune illnesses that target the brain, exemplified by multiple sclerosis. These advances reflect the impressive progress in human immunology and the range of new treatments generated as a result. However, a remarkable discovery over the past decade is the range of nervous system disorders that are not autoimmune but nevertheless involve neuroimmune mechanisms. A subset of depressed patients display a hyperinflammatory state,

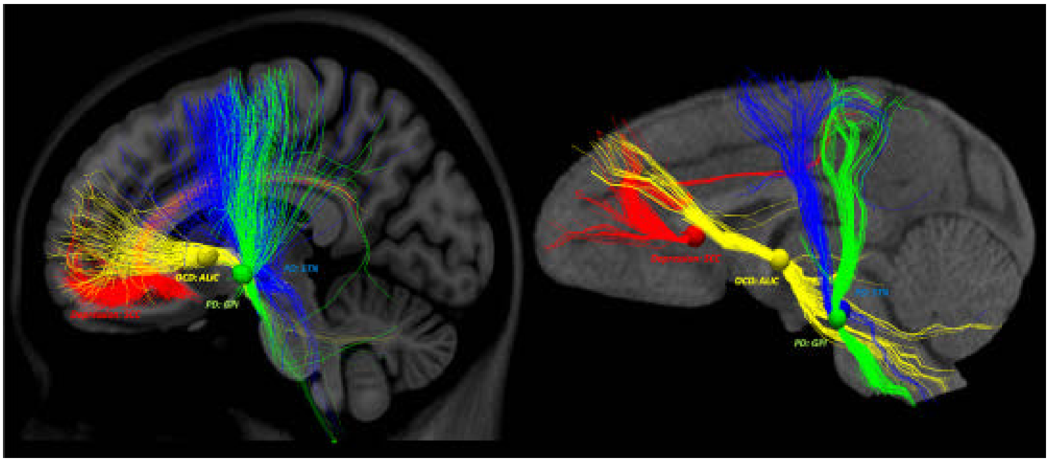


**FIGURE 2. Comparison of neural pathways affected by deep brain stimulation in human (left) and nonhuman primate (right) brain.** Magnetic resonance imaging (MRI) diffusion tractography illustrating the impact of deep brain stimulation in different neural circuits used to treat neuropsychiatric conditions including depression (red), obsessive compulsive disorder (yellow), and Parkinson's disease (green and blue). Image from Kisueng Choi, Juna Khang, Helen Mayberg, Peter Rudebeck, and Satoka Fujimoto (Icahn School of Medicine at Mount Sinai).

and the causal role of such neuroinflammation in controlling stress responses has been demonstrated in rodent models (*19*). These advances are guiding current clinical studies of anticytokine antibodies in treating this subset of depressed individuals. The importance of neuroimmune mechanisms is further underscored by the finding that some of the strongest known genetic risk factors for Alzheimer's disease are genes enriched in microglia—resident immune cells in the brain (*20*). Still another example, which may not relate to disturbed immune function per se, is the postulated involvement of complement pathways and microglia in disordered synaptic pruning that has been implicated in schizophrenia (*21*). The gut microbiome may contribute to a range of brain abnormalities in part by influencing peripheral immune mechanisms. Similarly, the interplay between cardiovascular health and brain function—which may be partly mediated by immune-based mechanisms controlling atherosclerosis and by brain circuits that ultimately influence atherosclerosis through the autonomic nervous system—are thought to explain the dramatic comorbidity between cardiovascular disease and brain disorders as diverse as stress-related disorders (depression, posttraumatic stress disorder); stroke; and Alzheimer's disease and other dementias (*22*). Furthermore, the autonomic nervous system influences all other peripheral organs and their diseases, an area of intense research today.

The control of intermediary metabolism is integrally related with brain function. A major achievement over the past two decades has been the elaboration of peripheral hormones (e.g., leptin, ghrelin, and insulin) that influence the brain, which in turn controls feeding behavior through circuitry that arises in the hypothalamus (*23*). How hypothalamic function then determines feeding behavior and energy expenditure beyond the autonomic nervous system remains a major challenge—reflecting the circuit–behavior divide discussed above. Interplays between the brain and peripheral metabolism are thought to explain the robust comorbidity seen for obesity and depression and implicated as well in addiction and many other brain syndromes (e.g., *24*).

## The ultimate frontier of modern medicine

The brain and nervous system will be the last frontier of medicine's precision revolution. This is expected given the unique complexity of the brain and its functions. Despite the challenges, however, the field is now at an inflection point where it is possible for the first time to delineate the myriad molecular constituents of individual cells, understand how those constituents determine cell function, and how the functioning of individual cells determine the activity of a single circuit and of larger and more complex composite circuits, ultimately producing behavior, then feeding back on cell and molecular function and so on. The new tools that study this bidirectional relationship across molecules, cells, and circuits are driving optimism that advances in understanding and treating brain disorders, which have long defied biomedical research, will be coming in the next decade. In fact, after decades of disappointment in therapeutics, the last few years alone have seen numerous examples of novel therapeutic mechanisms for psychiatric and neurological disorders, which are harbingers of far greater successes in the decade to come (e.g., *25*). It is impossible to overstate the positive impact on humanity that will be gained by one day conquering brain and nervous system disorders.

## References

1. S. K. Longo, M. G. Guo , A. L. Ji , P. A. Khavari, *Nat. Rev. Genet.* **22**, 627–644 (2021).
2. V. Kumar *et al.*, *Brain Struct. Funct.* **226**, 481–499 (2021).
3. J. F. Fullard *et al.*, *Genome Med.* **13**, 118 (2021).
4. B. M. Sweis, W. Mau, S. Rabinowitz, D. J. Cai, *Curr. Opin. Neurobiol.* **67**, 199–206 (2021).
5. P. Rajasethupathy, E. Ferenczi, K. Deisseroth, *Cell* **165**, 524–534 (2016).
6. B. L. Roth, *Neuron* **89**, 683–694 (2016).
7. V. P. Mathis, M. Williams, C. Fillinger, P. J. Kenny, "Networks of Habenula-Projecting Cortical Neurons Regulate Cocaine Seeking," *Sci. Adv.*, in press (2021).
8. M. G. Perich, K. Rajan, *Curr. Opin. Neurobiol.* **65**, 146–151 (2020).
9. Brainstorm Consortium *et al.*, *Science* **360**, eaap8757 (2018), doi: 10.1126/science.aap8757.
10. S. De Rubeis *et al.*, *Nature* **515**, 209–215 (2014).
11. E. J. Nestler, C. J. Peña, M. Kundakovic, A. Mitchell, S. Akbarian, *Neuroscientist* **22**, 447–463 (2016).
12. A. Bara, J. N. Ferland, G. Rompala, H. Szutorisz, Y. L. Hurd, *Nat. Rev. Neurosci.* **22**, 423–438 (2021).
13. Y. Y. Yim, C. D. Teague, E. J. Nestler, *Nat. Rev. Neurosci.* **21**, 471–484 (2020).
14. J. A. Kulkarni *et al.*, *Nat. Nanotechnol.* **16**, 630–643 (2021).
15. J. K. Krauss *et al.*, *Nat Rev Neurol.* **17**, 75-87 (2021).
16. P. E. Holtzheimer, H. S. Mayberg, *Annu. Rev. Neurosci.* **34**, 289–307 (2011).
17. H. G. Yi, M. K. Leonard, E. F. Chang, *Neuron* **102**, 1096–1110 (2019).
18. P. H. Rudebeck, E. L. Rich, H. S. Mayberg, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 26288–26296 (2019).
19. M. L. Pfau, C. Ménard, S. J. Russo, *Annu. Rev. Pharmacol. Toxicol.* **58**, 411–428 (2018).
20. A. Podleśny-Drabiniok, E. Marcora, A. M. Goate, *Trends Neurosci.* **43**, 965–979 (2020).
21. M. B. Johnson, B. Stevens, *Nature* **554**, 438–439 (2018).
22. K. Vandoorne *et al.*, *Circ. Res.* **123**, 415–427 (2018).
23. A. Moura-Assis, J. M. Friedman, L. A. Velloso. *Am. J. Physiol. Endocrinol. Metab.* **320**, E326–E332 (2021).
24. A. Duncan *et al.*, *Nature* **574**, 372–377 (2021).
25. J. H. Krystal, D. S. Charney, R. S. Duman, *Cell* **181**, 7 (2020).

PERSONALIZED MEDICINE

# The future of therapeutic discovery: Pioneering new pathways to personalized medicine

Paul J. Kenny[1,2]*, Nina Bhardwaj[3], Marta Filizola[4], Andrew Stewart[5]

Each person carries several million relatively common genetic variants that are widely distributed throughout the population, between 10,000 and 100,000 much less common variants shared with related family members, and around 100 de novo variants found exclusively in that individual (*1*). The unique complement of genetic variants we each carry, and how they interact with the different environments we inhabit, help to explain why some of us are more susceptible (or resilient) to certain diseases than others, and why the same drug treatment can be beneficial in one person yet be ineffective or even detrimental in another. The past decade has seen groundbreaking advances in our understanding of the basic mechanisms of disease, driven in part by breakthroughs in large-scale DNA sequencing (*2*). As genomes and transcriptomes are sequenced more frequently and with ever-greater cellular, spatial, and temporal resolution, so too will our understanding of the genetic architectures of disease continue to improve (Figure 1). Such progress will move medicine toward a future of personalized treatments designed by physicians to maximize an individual's therapeutic response while minimizing their risk of an adverse outcome.

The full potential of personalized medicine will only be realized as the arsenal of safe, effective treatment options available to physicians increases dramatically above current levels. A major hurdle impeding such progress is the exorbitant cost of developing new drugs, currently estimated at $2.6 billion per new-drug approval (*3*). The high failure rate of traditional small-molecule drug discovery is largely responsible for these costs. The reasons for such failures are manifold and include incomplete understanding of the drivers of disease; faulty assumptions about the involvement of a drug target in a disease; failure of a new drug candidate to engage its intended target; and detecting toxicity only after a new treatment is given to large patient populations. Finding creative solutions to these vexing problems will facilitate the future of therapeutic discovery.

The last two decades have seen dramatic improvements in computational approaches to drug discovery. Increased computational power and improved software platforms have made it possible to discover new compounds that modulate drug targets using in silico approaches. Computer-aided, structure-based drug discovery has enabled "virtual" screening of ultralarge libraries of chemically diverse compounds against drug targets at scales impossible to achieve using traditional "hands-on" approaches (*4*). Classical examples of successful computer-aided drug discovery include the human immunodeficiency virus (HIV) protease inhibitor amprenavir (*5*) and the antibiotic norfloxacin (*6*). This success was made possible by advances in X-ray crystallography, cryo-electron microscopy, and other biophysical techniques, which are yielding remarkably high-resolution, three-dimensional structures of drug targets and mechanistic information to support more accurate virtual screens.

The availability of expansive datasets of chemically, structurally, and biologically relevant data, in addition to faster computers and improved algorithms, has intensified the use of artificial intelligence (AI) and machine learning techniques for drug discovery (Figure 1) (*7*). Not only is AI helping to identify new uses for old drugs, but several companies have announced drug candidates designed de novo with the help of AI that have advanced to clinical trials (*8*). Even structural biology is benefiting from AI. The AlphaFold program developed by Google's DeepMind can generate molecular models of drug targets of crystallographic quality (*9*). Improving the ability of AI to identify new drugs will depend on a continuous supply of high-quality data to train reproducible machine learning models to accurately predict drug–target interactions and identify chemotypes with the most favorable drug-like properties. AI will also help address some of the other challenges that have plagued drug discovery efforts. These include more accurately predicting adverse drug side effects at the earliest stages of the discovery process, identifying the most tractable synthetic pathways to generate industrial quantities of drugs, and formulating novel synthetic pathways for compounds that otherwise cannot be generated in a commercially viable manner. Additionally, AI and machine learning may facilitate the discovery of drugs for protein or nonprotein targets that are considered undruggable. For example, there is much promise in modulating discrete disease-associated sequences or conformational states of DNA, and protein-coding or noncoding RNAs (*3*), yet our understanding of how genetic materials can be manipulated in a therapeutic manner is in its infancy.

Improvements in computer-aided drug design alone will not fix high failure rates in drug discovery. Better in vitro modeling of diseases will be required to define biological processes that drive disease, validate bona fide disease-modifying targets, and better predict drug candidates that are efficacious in patients. Innovations in patient-derived induced pluripotent stem cells (iPSCs), iPSC-derived organoids, and so-called "organs-on-a-chip" are yielding improved models of disease and cellular systems that are more capable of predicting adverse drug effects (*10*, *11*) (Figure 1). Continued advances in cell and tissue engineering, combined with improvements in materials science and microfabrication techniques, will further optimize the conditions under which patient-derived cells can better recapitulate healthy or pathological conditions in the human body (*12*).

Learning more about how environmental factors interact iteratively with an individual's genome to determine disease risk (Figure 1) (*13*) will catalyze the development of entirely new classes of prophylactic drugs, many of which are likely to operate through biological mechanisms fundamentally different from those targeted after a disease is established. The tremendous promise of genome-editing technologies may be realized such that disease-associated genetic variants are corrected before diseases occur (*2*). By editing RNA, it may be possible to manipulate transcriptional responses to disease-

---

[1]Nash Family Department of Neuroscience and [2]Drug Discovery Institute, [3]Department of Medicine and the Tisch Cancer Institute, [4]Department of Pharmacological Sciences, [5]Diabetes, Obesity, and Metabolism Institute, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: paul.kenny@mssm.edu

causing environmental insults so as to reduce disease risk or facilitate recovery. Progress in genetic and chemical engineering will be required to develop viruses, nanoparticles, lipid vesicles, modified retrotransposons, or other novel systems to deliver genetically encoded information, such as therapeutic transgenes or the proteins and RNAs required for genome editing, to only those cells in which expression is desired, and to do so in a safe, efficient, and temporally controlled manner (*14*, *15*).

While stunning progress has been achieved in sequencing and manipulating genetic materials, equivalent progress for proteins has been slower. This reflects the daunting technical hurdles associated with sequencing proteins and their posttranslational modifications in a high-throughput, accurate, and inexpensive manner. Progress in proteomics will better link disease-related genetic variants to the protein networks through which they operate and identify more robust biomarkers of disease vulnerability and therapeutic response to drug treatments. In parallel, novel therapeutic proteins with high selectivity for their targets and low risk of adverse events will complement or even supersede traditional small-molecule therapeutics for many diseases. Harnessing the power of the immune system will also play an increasingly important role in therapeutic discovery. Recombinant monoclonal, bispecific, monomeric, and drug-conjugated antibodies, as well as other immune proteins, are emerging as important interventions for many cancers (Figure 1). Genetic reengineering of immune cells, exemplified by chimeric antigen receptor (CAR)-T-cell-directed therapies, is likely to become more prevalent. Similarly, RNA vaccines, oncolytic viruses, checkpoint inhibitors, and other novel strategies to modulate the activity of the immune system and target immune cells to diseased tissue will continue to advance. The brain and other relatively immune-privileged organs represent new vistas for immunotherapeutics (*16*)

A final note on where drug development is likely to occur in the future: Recent years have seen major pharmaceutical companies reduce early-stage drug discovery efforts to

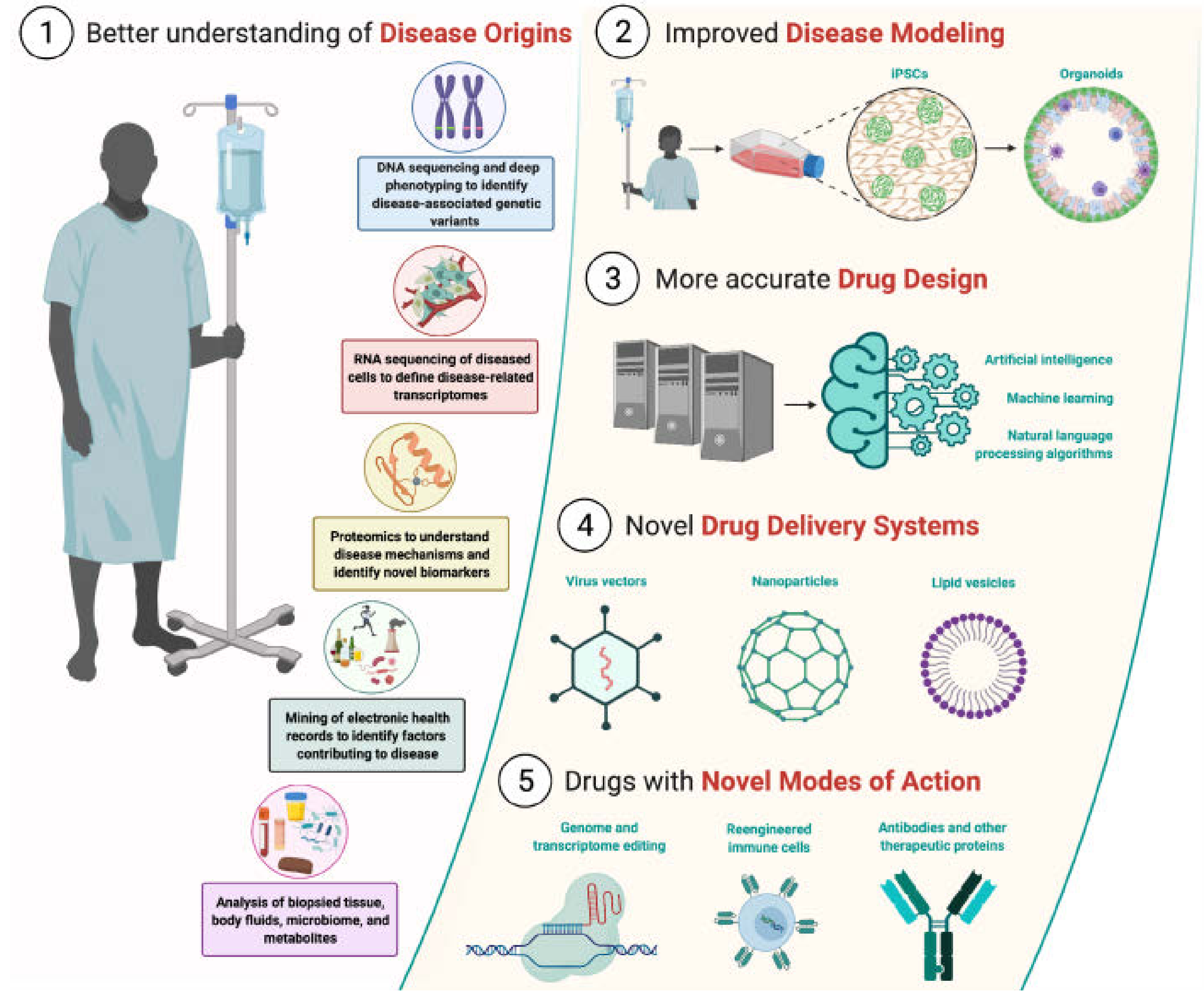

1 Better understanding of Disease Origins
DNA sequencing and deep phenotyping to identify disease-associated genetic variants
RNA sequencing of diseased cells to define disease-related transcriptomes
Proteomics to understand disease mechanisms and identify novel biomarkers
Mining of electronic health records to identify factors contributing to disease
Analysis of biopsied tissue, body fluids, microbiome, and metabolites
2 Improved Disease Modeling
iPSCs
Organoids
3 More accurate Drug Design
Artificial intelligence
Machine learning
Natural language processing algorithms
4 Novel Drug Delivery Systems
Virus vectors
Nanoparticles
Lipid vesicles
5 Drugs with Novel Modes of Action
Genome and transcriptome editing
Reengineered immune cells
Antibodies and other therapeutic proteins

**FIGURE 1. The future of therapeutic discovery.** Improvements in genomic and transcriptomic sequencing and proteomics technologies, and artificial intelligence (AI)-based mining of electronic medical records and digital data, will support the identification of new targets for therapeutic discovery. Advances in AI and other cutting-edge computational approaches will drive more efficient discovery of new drugs that are safe and effective. Cell and tissue engineering will deliver better in vitro models of disease. New therapeutic modalities will emerge and new classes of currently undruggable targets will become tractable. These advances will bring us closer to a future of truly personalized medicine iPSCs, induced pluripotent stem cells..

mitigate the high risk associated with these endeavors. Instead, companies are increasingly partnering with academic institutions and smaller biotech companies to fill their pipelines, often only after a novel therapeutic has been relatively derisked by successful advancement to an important preclinical or clinical development milestone. Governmental funding agencies have invested heavily in academic drug discovery initiatives over the past decade. In parallel, academic institutes have increasingly embraced the idea of translating their most promising new findings into drug discovery programs, a trend that is likely to continue as new drug successes are realized. This rapidly changing landscape in early-stage drug discovery presents unique challenges and opportunities for all involved. Finding new ground and fostering new types of relationships between academic institutes, pharmaceutical companies, and government agencies is critically important for the future success of therapeutic discovery.

## References

1. Institute of Medicine (U.S.), *Genome-Based Therapeutics: Targeted Drug Discovery and Development: Workshop Summary* (National Academies Press, Washington, DC, 2012).
2. E. E. Kenny, N. S. Abul-Husn, B. D. Brown, J. H. Cho, A. M. Goate, "Genetics and Gene Therapeutics: Igniting a Revolution in Health Care," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 16–18.
3. N. Fleming, *Nature* **557**, S55–S57 (2018).
4. M. Batool, B. Ahmad, S. Choi, *Int. J. Mol. Sci.* **20**, (2019).
5. A. Wlodawer, J. Vondrasek, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284 (1998).
6. A. C. Anderson, *Chem. Biol.* **10**, 787–797 (2003).
7. T. J. Fuchs et al., "The Advent of Artificial Intelligence in Health Care and its Practical Implementation," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 4–5.
8. N. Savage, *Biopharma Dealmakers* **15**, B37–B39 (2021), doi: 10.1038/d43747-021-00045-7.
9. J. Jumper *et al.*, *Nature* **596**, 583–589 (2021).
10. A. Sharma, S. Sances, M. J. Workman, C. N. Svendsen, *Cell Stem Cell* **26**, 309–329 (2020).
11. J. W. Blanchard et al., "Fulfilling the Promise of Stem Cell Biology and Regenerative Medicine," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 33–35.
12. A. Polinia, L. Moroniab, *Biomaterials and Biosystems* **1**, 100012 (2021).
13. R. O. Wright, K. N. Carroll, R. J. Wright, "Integrating Exposomics into Precision Medicine and Public Health," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 14–16.
14. K. Bulaklak, C. A. Gersbach, *Nat. Commun.* **11**, 5820 (2020).
15. M. Segel *et al.*, *Science* **373**, 882–889 (2021).
16. M. Merad *et al.*, "Building the Future of Immunology in Medicine," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 21–23.

---

**STEM CELL BIOLOGY AND REGENERATIVE MEDICINE**

# Fulfilling the promise of stem cell biology and regenerative medicine

**Joel W. Blanchard[1,2,3†], Timothy A. Blenkinsop[1,4,5 †], Saghi Ghaffari[1,4,6,7†], Robert S. Krauss[1,4,6,7†], Michael Rendl[1,4,8†], Christopher M. Sturgeon[1,4,9†], Thomas P. Zwaka[1,4†], Marek Mlodzik[1,4,5,6,7*], Sarah E. Millar[1,4,6,8*]**

Stem cell biology and regenerative medicine have long held promise for the development of new therapeutic strategies. A congruence of recent discoveries and innovations has propelled the field to the cusp of realizing this promise (Figure 1). These include the ability to generate autologous (patient-specific) human induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs) engineered to serve as a universal source for allogeneic cell therapy; the advent of gene-editing technologies to repair or introduce disease-associated mutations for personalized cell-based therapies or mechanistic studies; stem cell differentiation strategies based on information from developmental and single-cell analyses; and the use of bioengineering technologies to generate three-dimensional "mini-organs." Applications of stem cell biology include transplantation of stem cell–derived therapeutic cells or tissues; generation of in vitro platforms for mechanistic investigations and drug screening; and studies of endogenous stem cells and their niche environment that reveal new therapeutic targets.

Stem cell–based therapies are most immediately applicable in scenarios where cell transplantation procedures are already established in clinical practice. The classic example of bone marrow transplantation (BMT) relies on hematopoietic stem cells (HSCs) that give rise to all blood cells. BMT is a curative treatment for blood cancers and immune cell deficiencies; however, it relies on matched donors for allogeneic transplantation (*1*), and HSCs are short-lived ex vivo. These challenges are being addressed by developing new techniques to expand HSCs while maintaining potency (*2*), including altering the degradation capacity of lysosomes in HSCs (*3*). A fascinating but yet-to-be-demonstrated approach is to generate HSCs via directed differentiation of hiPSCs (*4*). This could potentially be scaled for banking of common haplotypes for transplant, or employed in a patient-specific setting with gene editing of autologous iPSCs to correct gene defects in inherited diseases such as sickle cell anemia.

Another exciting application of hPSC differentiation technologies is the generation of immune cells for adoptive immunotherapies, for instance chimeric antigen receptor (CAR) T cells that are engineered to target tumors and can be produced at scale. T cells and natural killer cells can be stably engineered to have "off-the-shelf" anticancer potential and do not require the derivation of an HSC precursor for their generation from hPSCs. This in turn raises entirely new paradigms for hematopoeisis in regenerative medicine, as hPSCs robustly give rise to HSC-independent embryonic hematopoietic progenitors that may have unique therapeutic potential. For example, a recently identified embryonic myeloid progenitor gives rise to natural killer cells with enhanced cytotoxicity as compared with their HSC-derived counterparts (*5*). These cells are being tested as potential cancer therapeutics. In parallel, gene-edited iPSC-derived myeloid cells are being used as a model to uncover mechanisms of malignant and nonmalignant blood diseases and to guide new therapies (*6*).

Other stem cell transplantation paradigms include the skin epidermis and cornea, which can respectively be regenerated by grafting of epidermal and corneal limbal stem cells (*7*). In one encouraging study, a therapeutic gene was virally

---

[1] Black Family Stem Cell Institute; [2] Ronald M. Loeb Center for Alzheimer's Disease; [3] Department of Neuroscience; [4] Department of Cell, Developmental and Regenerative Biology; [5] Department of Ophthalmology; [6] Tisch Cancer Institute; [7] Department of Oncological Sciences; [8] Department of Dermatology; [9] Division of Hematology and Medical Oncology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY

[†] These authors contributed equally.

*Corresponding authors: marek.mlodzik@mssm.edu (M.M.); sarah.millar@mssm.edu (S.M.)

introduced into epidermal stem cells that were expanded and grafted to cure an otherwise lethal genetic skin condition (*7*). More recently, several groups have used gene editing to correct the gene mutated in recessive epidermolysis bullosa in patient-derived iPSCs, followed by iPSC differentiation into keratinocytes for skin grafting (*8*). Retinal pigmented epithelium (RPE) is likewise amenable to stem cell transplantation to potentially treat vision loss. For example, RPE engineered from adult stem cells can safely replace host RPE function in nonhuman primates (*9*). hiPSC-derived RPE has been transplanted into a human patient, but this was not associated with a significant improvement in visual acuity, and a second patient was not treated due to genomic instability of the hiPSC-derived cells (*10*). Other concerns surrounding hiPSC-based therapies include the preservation of epigenetic memory in hiPSC-derived cells that may impact their differentiation potential in vivo. Thus, these approaches still face significant hurdles to their adoption in clinical practice.

Researchers have harnessed the self-assembling properties of differentiating stem cells to partially recapitulate the development of more complex human organs, such as lung, small intestine, kidney, and brain to an extent (*11*, *12*), ex vivo. hiPSC-based brain organoids are being developed to recreate the choroid plexus, myelination, and the blood–brain barrier (*13*, *14*). Some models have even produced evidence of interaction of hiPSC-derived human brain tissue with other organs such as the gut-brain axis and neuromuscular circuits (*15*, *16*). Organoids are facilitating discoveries into fundamental disease mechanisms in many areas, including neurodevelopmental disorders (*17*), neurodegenerative diseases (*18*), and viral infection (*19*), and provide an important counterpart to animal models that can fail to recapitulate human disease accurately or may respond differently to therapeutic agents. However, their utility for disease modeling and drug screening relies on the extent to which they accurately reflect human physiology. For instance, comparative analysis of primary human fetal tissue and brain organoids at the single-cell level revealed important differences in the presence, maturation, and spatial distribution of distinct cell subtypes. Improving these models will require a deeper understanding of human fetal development, an area that has been hampered by restrictions on the use of discarded human fetal tissue in research (*20*).

Despite differences in detailed mechanisms and cell types between human and mouse model systems, the latter continue to play a key role in regenerative medicine. For instance, while grafting epidermal stem cells effectively replaces the outer layer of skin, such grafts lack hair follicles, oil glands, and touch cells. Current research using mouse genetic models is uncovering signaling pathways that direct development of these associated structures (*21*, *22*). In parallel, understanding the regulatory

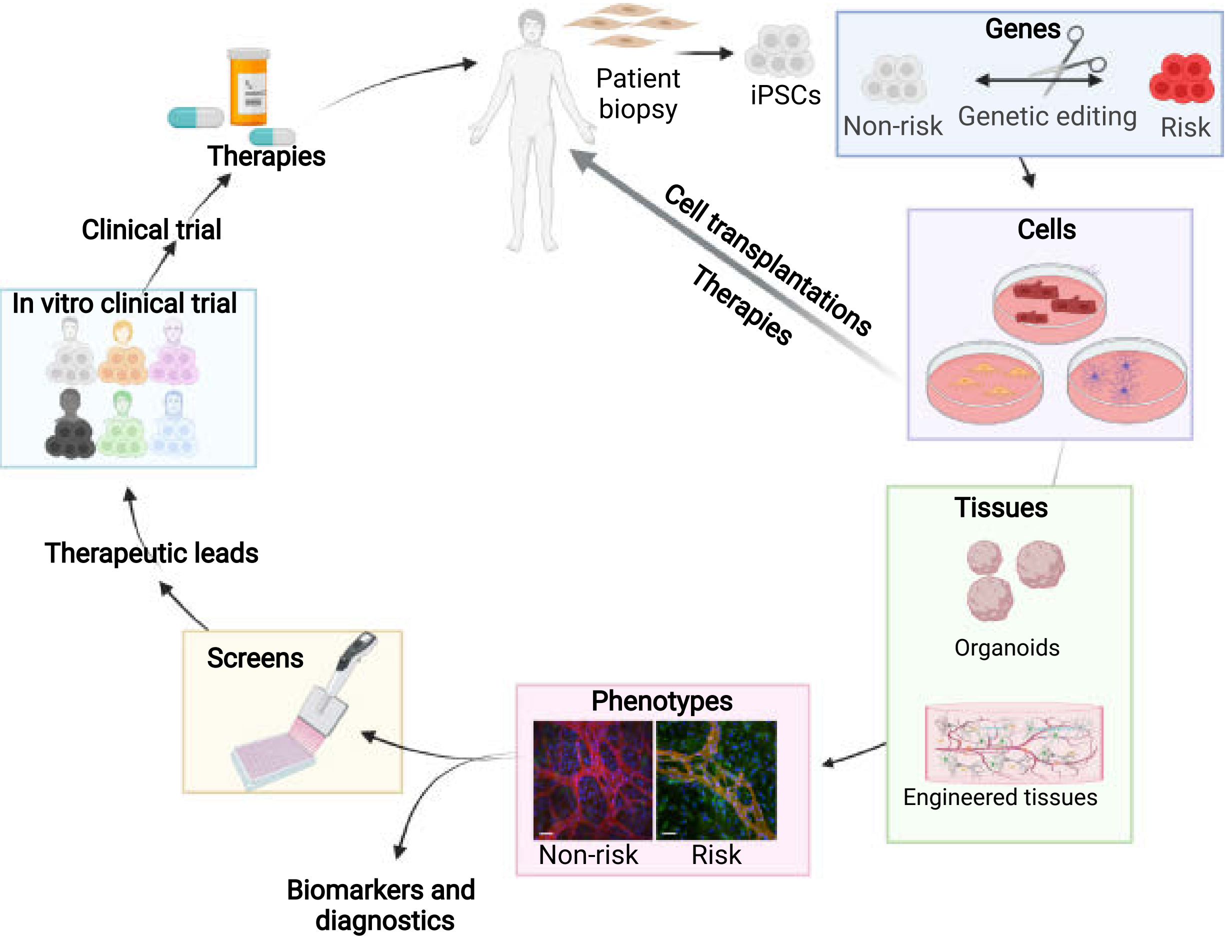


**FIGURE 1. Stem cell technology driving pathophysiological and therapeutic discovery.** iPSCs, induced pluripotent stem cells.

interactions between endogenous tissue stem cells and their niche environment suggests approaches to therapeutically stimulate or dampen stem cell activity. As an example, muscle wasting diseases have devastating effects and are currently incurable. Experiments in mice have shown that skeletal-muscle regeneration relies on muscle stem cells (MuSCs), which are normally maintained in a quiescent state by signals from the niche environment but are activated upon injury (*23–25*). While widespread therapeutic transplantation of MuSCs would clearly be challenging, alternatives may include use of biologicals or small molecules that target niche signals to enhance the function of native MuSCs.

This is a highly exciting era for regenerative medicine. While the next decade will undoubtedly reveal additional challenges related to safety and efficacy, we can expect to see the actualization of novel therapies based on multiple aspects of stem cell biology, particularly in areas where cell and tissue transplantation techniques are well established.

## References

1. J.-M. Tiercy, *Haematologica* **101**, 680–687 (2016).
2. E. Zimran, L. Papa, R. Hoffman, *Blood Rev.*, 100853 (2021).
3. R. Liang *et al.*, *Cell Stem Cell* **26**, 359–376.e7 (2020).
4. A. Fidanza, L. M. Forrester, *J. Immunol. Regen. Med.* **13**, 100050 (2021).
5. C. Dege *et al.*, *Dev. Cell* **53**, 229–239.e7 (2020).
6. T. Wang *et al.*, *Cell Stem Cell* **28**, 1074–1089.e7 (2021).
7. T. Hirsch *et al.*, *Nature* **551**, 327–332 (2017).
8. S. Choudhury, N. Surendran, A. Das, *Wound Repair Regen.* **29**, 697–710 (2021).
9. Z. Liu *et al.*, *Stem Cell Reports* **16**, 237–251 (2021).
10. M. Mandai *et al.*, *N. Engl. J. Med.* **376**, 1038–1046 (2017).
11. F. Schutgens, H. Clevers, *Annu. Rev. Pathol.* **15**, 211–234 (2020).
12. M. A. Lancaster *et al.*, *Nature* **501**, 373–379 (2013).
13. J. W. Blanchard *et al.*, *Nat. Med.* **26**, 952–963 (2020).
14. L. Pellegrini *et al.*, *Science* **369**, eaaz5626 (2020).
15. J. Andersen *et al.*, *Cell* **183**, 1913-1929.e26 (2020).
16. M. Trapecar *et al.*, *Sci. Adv.* **7**, eabd1707 (2021).
17. A. Tian, J. Muffat, Y. Li, *J. Neurosci.* **40**, 1186–1193 (2020).
18. K. Leng *et al.*, *Nat. Neurosci.* **24**, 276–287 (2021).
19. A. Z. Eriksen *et al.*, *Cell Stem Cell* **28**, 1205–1220.e7 (2021).
20. M. Haniffa *et al.*, *Nature* **597**, 196–205 (2021).
21. K. W. Mok *et al.*, *Dev. Cell* **48**, 32–48.e5 (2019).
22. I. Cohen *et al.*, *Cell Stem Cell* **22**, 726–739.e7 (2018).
23. A. P. Kann, M. Hung, R. S. Krauss, *Curr. Opin. Cell Biol.* **73**, 78–83 (2021).
24. S. Eliazer *et al.*, *Cell Stem Cell* **25**, 654–665.e4 (2019).
25. A. J. Goel, M. K. Rieder, H. H. Arnold, G. L. Radice, R. S. Krauss, *Cell Rep.* **21**, 2236–2250 (2017).

---

**SURGERY**

# Dramatic innovations in surgery

**Joshua Bederson[1]*, David H. Adams[2], Evan Flatow[3], Leesa Galatz[3], Eric M. Genden[4], Michael L. Marin[5]**

Although surgical training is long and rigorous, even the most creative surgeons are limited by human biology and available tools. To cite one example, blood is red, tissues are opaque, and we cannot see through the body. Technological solutions to these and myriad other limitations have driven a revolution in surgical innovation over the past 50 years. Advances in microscopy, miniaturization, implantable materials, imaging techniques, and our understanding of biology and brain function, together with powerful computing, robotics, and artificial intelligence (AI), have led to dramatic changes in the way surgery is performed and to the development of entirely new treatments.

### Orthopedics and spine surgery

An explosion of basic science tools and the demands of an increasingly aging and yet active population have amplified the need for less invasive and regenerative approaches. Percutaneous fracture repair and new ingrowth materials for joint replacement (*1*) have been developed. Computer-aided navigation, robotics, patient-specific instrumentation, and augmented/virtual reality during spine and joint replacement surgery have increased safety (*2*) and have enabled surgeons to save joints that were previously considered inoperable due to bone loss in degenerative arthropathies. Research into the biology of cartilage and tendon (*3*) and how they repair after injury is driving development of tissue engineering and cell therapies to enhance healing (*4*).

### Vascular and cerebrovascular surgery

Treatment of the most common and devastating diseases, including vascular insufficiency, stroke, and aneurysms, has been completely revolutionized within the last decade. Flow-diverting and fenestrated endovascular prostheses allow reconstruction of diseased arteries, resulting in the cure of previously untreatable aneurysms of the brain (*5*) and aorta (*6*). Intracranial thrombectomy is the most effective medical treatment to date (*7*). Endovascular access has opened a pathway to treat disorders outside vasculature, such as in cardiac arrhythmias and now in brain disorders (Figure 1).

### Tumor surgery

Maximal resection while preserving function is the goal for tumor surgery. However, many tumors are not visible to the naked eye or even through an operating microscope. Frameless stereotactic navigation has become a modern standard in brain tumor surgery (*8*), and combined with other techniques, greatly enhances the ability to differentiate tumor from normal tissue. New technologies include intraoperative magnetic resonance imaging (MRI), ultrasound, and electrophysiological brain mapping. New molecular agents aid in selectively visualizing tumor cells (*9*), and some of the same fluorophores that delineate tumor cells may be used to eradicate them by light activation. Use of magnetic nanoparticles (*10*), patient-specific

---

[1]Department of Neurosurgery, [2]Department of Cardiovascular Surgery, [3]Department of Orthopedics, [4]Department of Otolaryngology, [5]Department of Surgery, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: joshua.bederson@mountsinai.org

induced pluripotent stem cells (*11*), and low-intensity focused ultrasound are all being developed to enhance tumor resection.

### Cardiac surgery

Deep collaborations between cardiac surgeons, electrophysiologists, and cardiac imaging scientists have led to new concepts in the management of heart diseases such as mitral valve prolapse (MVP), which affects up to 3% of the Western population. MVP usually follows a protracted course, but some patients suffer devastating ventricular arrhythmias or fatal sudden cardiac arrest early in the disease. By combining electrophysiology, positron emission tomography, and MRI, a previously unrecognized link was found between structural abnormalities in the mitral valve and the deadly arrhythmias (Figure 2). Direct-access cryoablation of a pathological focus at the time of mitral repair surgery led to a dramatic reduction in abnormal activity. This work has inspired an emerging national focus on the treatment of patients with associated ventricular arrhythmias (*12*, *13*).

### Augmented and virtual reality

Advances in neuroimaging and computer graphics have permitted the ability to merge different modalities such as computerized tomography (CT), MRI, angiography, and ultrasound to create interactive three-dimensional (3D) virtual reality "scenarios." These can be coregistered to a patient's anatomy and linked to navigation devices that are essentially GPS maps of organs—such as the human brain—that guide surgery. Projected into an operating microscope in a heads-up display during surgery, these systems are at least a partial solution to a vexing problem, allowing surgeons to see "through" tissues to the critical structures beyond (*14*) (Figure 3).

### Restorative surgery

Even the most advanced instruments may not be capable of restoring function in patients with trauma, cancer, stroke, or neurodegenerative disorders. A deeper understanding of neural stem cell physiology and 3D brain organoids, together with novel transplant strategies, has led to enhanced engraftment of transplanted brain tissues (*15*). Improved understanding of axonal regeneration and wound healing has pushed the development of novel regenerative approaches that encourage neural repair and axonal reconnectivity after traumatic injury (*16*). Advanced immunosuppressive methods combined with creative surgical techniques have led to the first-in-human transplantation of a trachea and adjacent structures (*17*). These advances have ushered in a new era of transplantation that promises to restore complex functions such as movement after paralysis, control of tremors, and the return of smell, taste, speech, and swallowing.

### Brain–computer interfaces

From the first reports of implantable brain–computer interfaces (BCI) to the modern era, BCI can be seen as developing along at least two pathways. In one, electrodes, stimulators, and programmable brain pacemakers modulate, suppress, or block pathological processes such as epilepsy, movement disorders such as Parkinson's disease (*18*), and

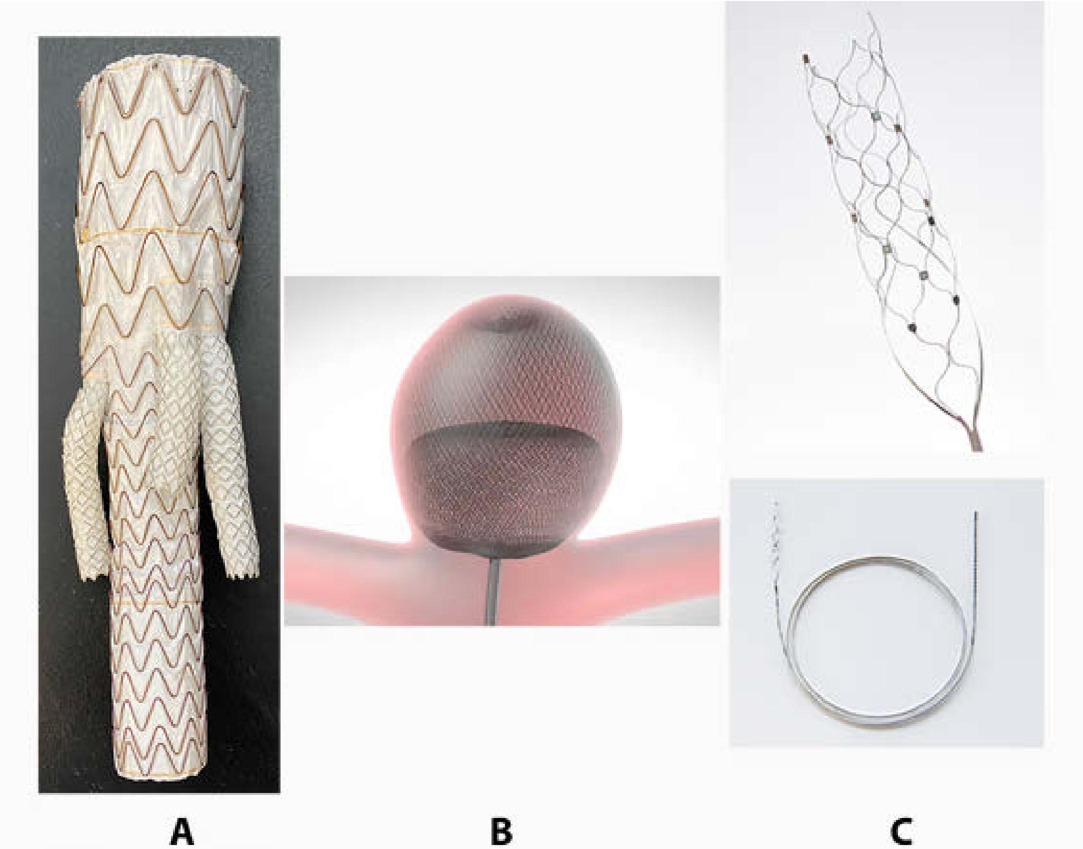


**FIGURE 1. Examples of modern endovascular devices.** (**A**) Aortic endovascular device with custom side branches. Scale is approximately 4 x 18 cm. (**B**) Intracranial flow-diverting stent used to reconstruct brain arteries with a diameter of approximately 3 mm. (**C**) "Stentrode" with sixteen 500 μm-diameter recording electrodes embedded in an intracranial stent and connected to a subclavicular transmitting device (see Figure 4).

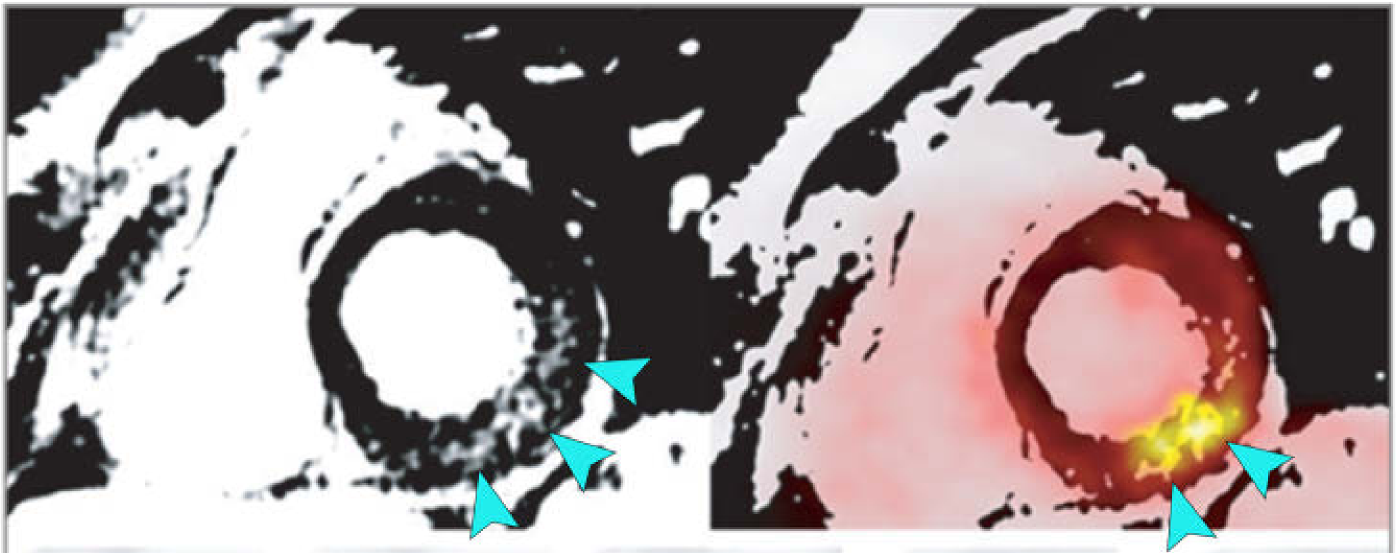


**FIGURE 2. Hybrid positron emission tomography/magnetic resonance imaging (PET/MRI) of an asymptomatic patient.** Example of hybrid cardiac PET/MRI demonstrating subclinical subendocardial injury in an apparently normal ventricle by routine standard echo: fluorine 18–labeled fluorodeoxyglucose uptake (inflammation marker; yellow color, **right panel**) and late gadolinium enhancement (scar/fibrosis marker; light grey, **left panel**). Asymptomatic patient with mitral valve prolapse causing significant mitral regurgitation and complex ventricular ectopy. Adapted from (*12*).

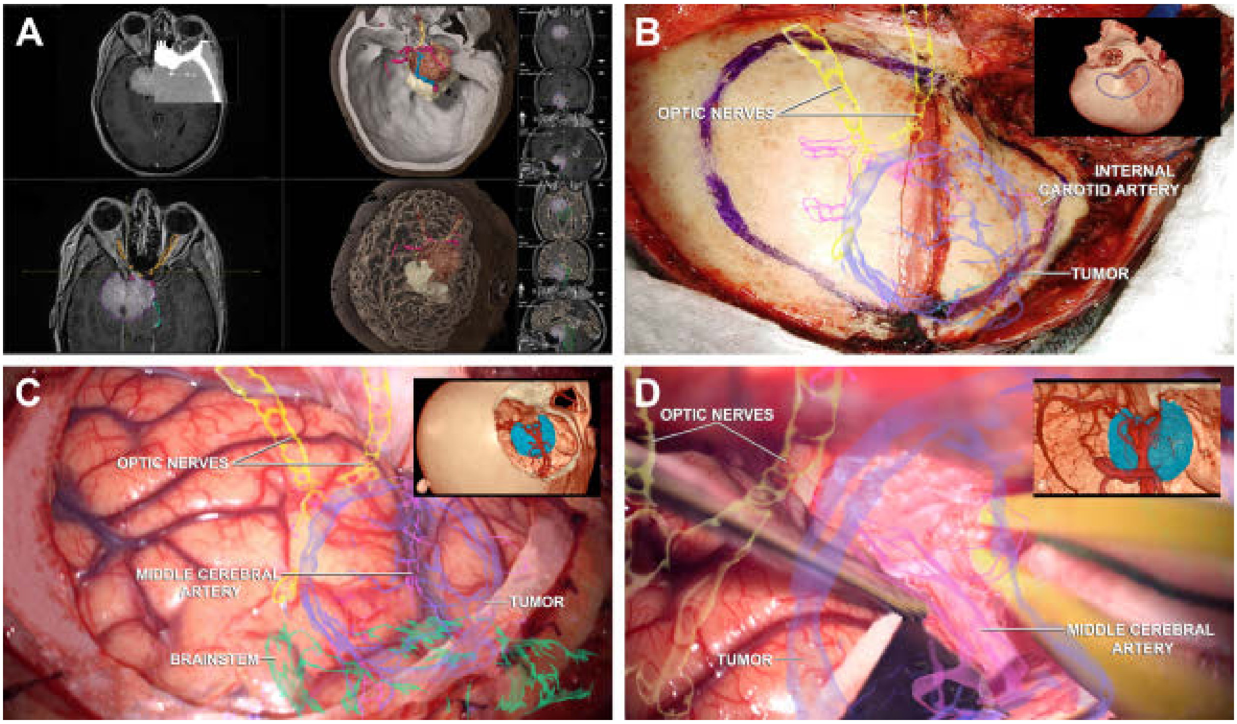


**FIGURE 3. Use of navigation-linked augmented reality and heads-up display in brain tumor removal.** (**A**) Fusion of computed tomography angiography and magnetic resonance imaging; selection of brain structures; creation of 3D model. (**B**) Projection of selected objects on skull during craniotomy. Inset shows virtual reality model. (**C**). After craniotomy, the critical structures are visible "through" the brain during exposure. (**D**) Under higher magnification, the critical structures and tumor borders are seen in anticipation of their exposure.

neuropsychiatric conditions such as obsessive-compulsive disorder, Tourette syndrome, and depression (*19*). Full closed-loop self-programming systems detect abnormal signaling patterns and adjust the output to suppress pathological activity in the brain (*20*). As our understanding of brain circuitry improves, dementia, traumatic brain injury, addiction, obesity, and even cognitive enhancement (*21*) are potential indications for modulation.

In another major pathway, BCI technologies aim to address failed transmission from the cerebral cortex due to conditions such as spinal cord injury, stroke, and amyotrophic lateral sclerosis. Surgically implanted neuroprostheses create a direct link from the brain to computers. For example, data from cortical surface electrodes have been used to produce natural speech (*22*), and using AI algorithms, to decode handwriting translated to texting. One such invention is the Stentrode, an endovascular stent and brain–machine interface already implanted in four humans and helping to translate their thoughts into digital output (Figure 4) (*23*). It is essential to develop tactile or haptic feedback perception using BCI; until this occurs, even a high-performing motor solution will be perceived as separate or unwieldy by its user.

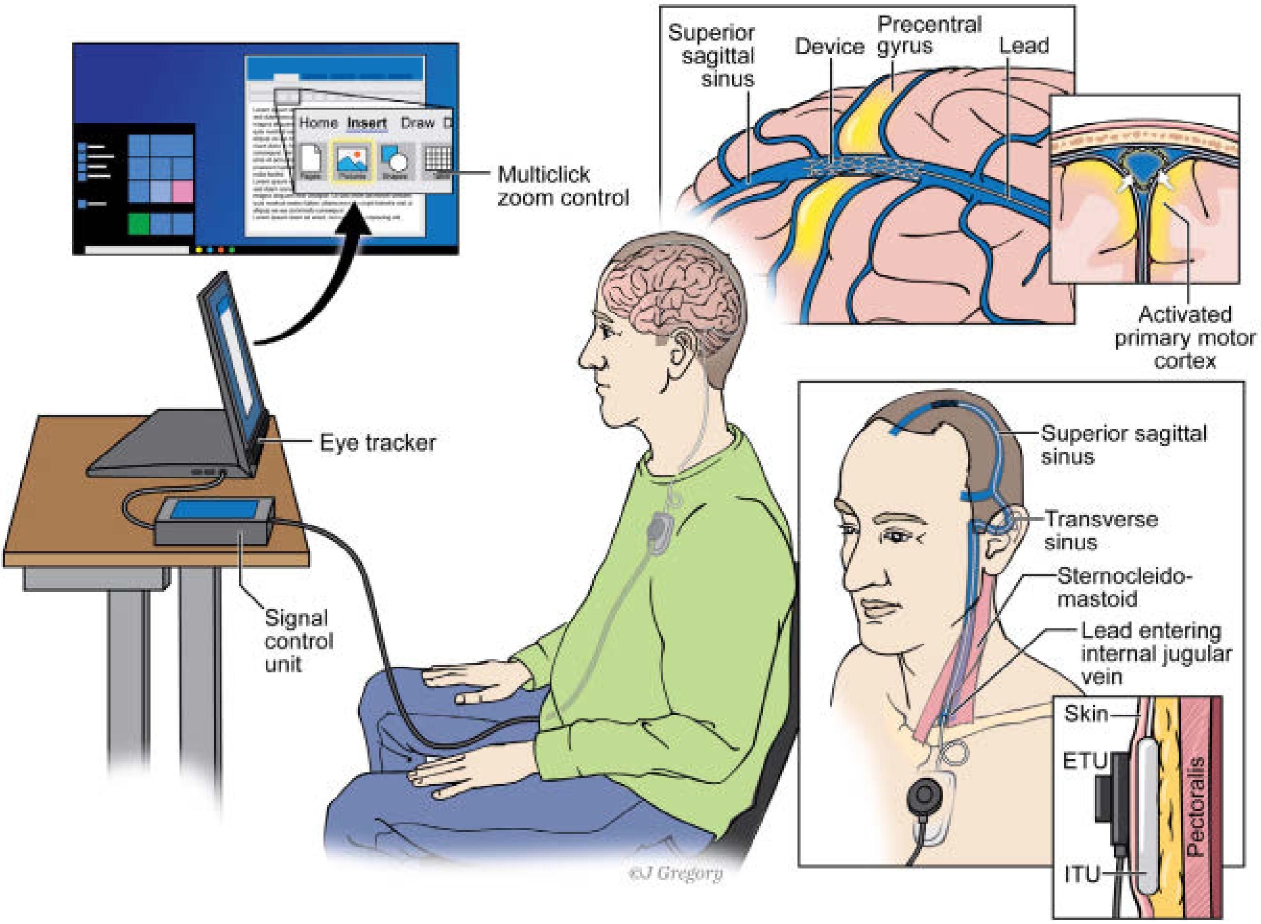


**FIGURE 4. Computer control via an endovascular motor neuroprosthesis in a patient with flaccid upper-extremity paralysis.** This participant had complete flaccid upper-limb paralysis due to motor neuron disease. The endovascular Stentrode was implanted in the superior sagittal sinus adjacent to the precentral gyrus (highlighted yellow region), which is activated with attempted limb movement. The transmission lead is connected to a telemetry unit within a subclavicular pocket, and via infrared light transmission, sends the electrocorticography signal to a computer. The signals are translated into multiple click-and-go actions, which are combined with eye-tracking to enable the general operation of a Microsoft Windows 10 interface. ETU, external telemetry unit; ITU, internal telemetry unit.

Commercialization of BCI technology is an important driver in the field, harnessing the massive resources available to industry. Likewise, continuous reevaluations of moral principles by physicians and scientists will be critical in guiding future research and care in this rapidly advancing field (*24*).

## Access to care

Finally, technological innovations coupled with improved at-home care are expanding the platform for ambulatory surgery, which can be applied to the 5 billion people in the world who currently do not have access to surgical interventions (*25*).

## Acknowledgments

The authors would like to thank Laura Salgado-Lopez, Benjamin Rapoport, Dimosthenis Pandis, Christopher Kellner, Helen Mayberg, Hongyan Zou, Costas Hadjipanayis, Isabelle Germano, Saadi Ghatan, Jeremy Steinberger, Ted Panov, Thomas Oxley, and Brian Kopell for thoughtful comments and contributions to this work.

## References

1. A. K. Harrison *et al.*, *J. Bone Joint Surg. Am.* **94**, 1223–1228 (2012).
2. J. R. Joseph *et al.*, *Neurosurg. Focus* **42**, E2 (2017).
3. F.-J. Lyu *et al.*, *Bone Res.* **9**, 7 (2021).
4. L. M. Galatz *et al.*, *J. Orthop. Res.* **33**, 823–831 (2015).
5. T. Becske *et al.*, *Radiology* **267**, 858–868 (2013).
6. G. S. Oderich *et al.*, *J. Vasc. Surg.* **70**, 358–368.e6 (2019).
7. M. Goyal *et al.*, *N. Engl. J. Med.* **372**, 1019–1030 (2015).
8. M. A. Kesserwan *et al.*, *Surg. Neurol. Int.* **12**, 52 (2021).
9. W. Stummer *et al.*, *Lancet Oncol.* **7**, 392–401 (2006).
10. G. P. Skandalakis *et al.*, *Int. J. Hyperthermia* **37**, 3–19 (2020).
11. E. Mormone *et al.*, *Stem Cells Dev.* **23**, 2626–2636 (2014).
12. M. A. Miller *et al.*, *JAMA Cardiol.* **5**, 1000–1005 (2020).
13. A. El-Eshmawi *et al.*, *J. Am. Coll. Cardiol.* **76**, 3061–3062 (2020).
14. J. R. Mascitelli *et al.*, *Oper. Neurosurg. (Hagerstown)* **15**, 184–193 (2018).
15. N. Daviaud, R. H. Friedel, H. Zou, *eNeuro* **5**, ENEURO.0219-18.2018 (2018).
16. X. Zhou *et al.*, *Nat. Neurosci.* **23**, 337–350 (2020).
17. E. M. Genden *et al.*, *Am. J. Transplant.* **21**, 3421–3427 (2021).
18. P. Limousin *et al.*, *N. Engl. J. Med.* **339**, 1105–1111 (1998).
19. H. S. Mayberg *et al.*, *Neuron* **45**, 651–660 (2005).
20. D. R. Nair *et al.*, *Neurology* **95**, e1244–e1256 (2020).
21. C. S. Inman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 98–103 (2018).
22. D. A. Moses *et al.*, *N. Engl. J. Med.* **385**, 217–227 (2021).
23. T. J. Oxley *et al.*, *J. Neurointerv. Surg.* **13**, 102–108 (2021).
24. M. Hegde, W. Chiong, V. R. Rao, *Neurology* **96**, 799–804 (2021).
25. M. L. Marin, "Improving Access to Health Care from 7,000 Miles Away, with Michael Marin, MD," *Connected and Ready* podcast (July 14, 2021); available at https://connectedandready.com/episodes/improving-access-to-health-care-from-7-000-miles-away-with-michael-marin-md.

**WOMEN'S HEALTH**

# Comprehensive approaches to sex and gender science

Leslee J. Shaw[1,2,3,4]*, Veerle Bergink[1,5], Joanne L. Stone[4], Stephanie V. Blank,[4] Annetine C. Gelijns[2], Roxana Mehran[3]

Sex and gender differences in disease manifestation, clinical outcomes, and treatment effectiveness are understudied but full of enormous possibilities for discovery and advancement in women's health. It is only recently that research has focused on uncovering unique attributes that define sex-based phenotypes, including the many trait characteristics of women of diverse racial and ethnic backgrounds. This coincides with the U.S. National Institutes of Health's (NIH) 2016 mandate to include sex as a biological variable in all funded research, which at long last provides foundational knowledge of sex differences in cell and organ function in health and disease. Even though most research focuses on episodic care for a specific health condition, each life stage of a woman, from childhood to elder years, is inextricably linked with detectable precursors of future health care needs and chronic diseases in later life. Altering the pathways that lead to common chronic conditions, such as diabetes, hypertension, cancer, cardiovascular disease, and depression, requires pioneering research bringing together scientists and disease specialists to reveal a comprehensive understanding of health and disease in women (Figure 1).

Pregnancy-related complications are increasingly common, and—as our research underscores—highlight the alarming and increasing inequities of care for Black as compared to White women (*1*). Certain pathogenic conditions during pregnancy are precursors for a woman's later health, such as pre-eclampsia and future risk of hypertension. An array of pregnancy-related complications predicts future cardiovascular disease (such as hypertensive disorders of pregnancy, preterm birth, fetal growth restriction, gestational diabetes, and placental abruption). A model of early intervention has the potential to improve perinatal outcomes while also resulting in the immeasurable benefit of improving the health of middle–older aged women. However, research has lagged, and randomized trials evaluating the clinical benefit of intensified preventive care strategies aimed to reduce adverse perinatal outcomes are long overdue. Redefining high-quality care would importantly benefit women at highest risk, including those of diverse race and ethnicity. Mental health benefits would also be realized for diverse women, as health discrimination increases stress, anxiety, and depression during pregnancy (*2*). The mental health stability of mothers is also vital for the health of their children, as high stress levels or exposure to environmental toxins during pregnancy increase risk of mental health and cognitive problems in their offspring. Postpartum pathology is prevalent, notably among Black women, where infectious mechanisms drive not only cardiomyopathy but also co-occur with high rates of anxiety and depression. Targeted research aimed at improving preventive health practices would not only reduce near-term events related to pregnancy but dramatically reduce future risk of chronic diseases that greatly impact and debilitate a woman's later years, such as diabetes, hypertension, cancer, mood disorders, and cardiovascular disease.

As a woman ages, the loss of estrogen that occurs during menopause is associated with wide ranging disease expression and the emergence of several chronic diseases (*3*–*5*). Cardiovascular disease is a prominent condition, which increases in prevalence and severity in the postmenopausal woman. Despite marked advances in cardiovascular care, significant disparities in care persist, which adversely impact morbid and fatal outcomes for millions of women. Cardiovascular disease in women remains understudied, underrecognized, underdiagnosed, and undertreated. Ischemia with nonobstructive coronary artery disease, myocardial infarction with normal/nonobstructive coronary arteries, and spontaneous coronary artery dissection all commonly occur in women without a definable diagnostic and therapeutic strategy of care. Recently, *The Lancet* Commission on Women and Cardiovascular Disease addressed these worrisome trends by identifying knowledge gaps in research, prevention, treatment, and access to care for women (*6*). A wide range of psychological, social, economic, cultural, and sex-specific risk factors need further investigation and consideration to improve our understanding of cardiovascular disease in women.

Cross-disciplinary explorations focusing on women undergoing premature menopause is a means for early detection of cardiovascular disease. Early menopause before 40 years of age occurs more often among women who are obese, smokers, human immunodeficiency virus–infected, and those with certain autoimmune diseases. Iatrogenic premature loss of ovarian function is increasingly common for women at elevated genetic risk of ovarian cancer, such as those harboring pathogenic *BRCA1* or *BRCA2* variants who undergo risk-reducing bilateral salpingo-oophorectomy (*7*). Knowledge can also be gleaned through clinical and translational research in transgender/gender nonbinary people in whom feminizing or masculinizing hormone therapy may promote or attenuate atherosclerotic plaque progression.

The unique role of inflammation in women warrants further investigation, with autoimmune diseases occurring at a rate twofold higher than that of men (*8*, *9*). Population evidence reveals that women live within a greater inflammatory milieu, with higher levels of the generalized inflammatory marker C-reactive protein across all age groups (*10*–*12*). The high rate of obesity in women may contribute to this chronic, low-grade inflammation, and the associated increased inflammatory cytokine production may in turn contribute to obesity-related pathologies, such as diabetes, cancer, depression, and cardiovascular disease (*13*). Moreover, the post-COVID-19 syndrome occurs more often in women versus men (ratio=4:1), necessitating research unearthing the mechanism for the stronger autoimmune response among women (especially those younger women with potential underlying autoimmune disorders).

Disentangling the interplay between sex-specific traits and cultural and societal disadvantages impacting women's health is daunting. Until now, typical sex-based comparisons were

---

[1]Blavatnik Family Women's Health Research Institute, [2]Department of Population Health Science and Policy, [3]Department of Medicine, Division of Cardiology, [4]Department of Obstetrics, Gynecology, and Reproductive Science, [5]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY

*Corresponding author: leslee.shaw@mountsinai.org

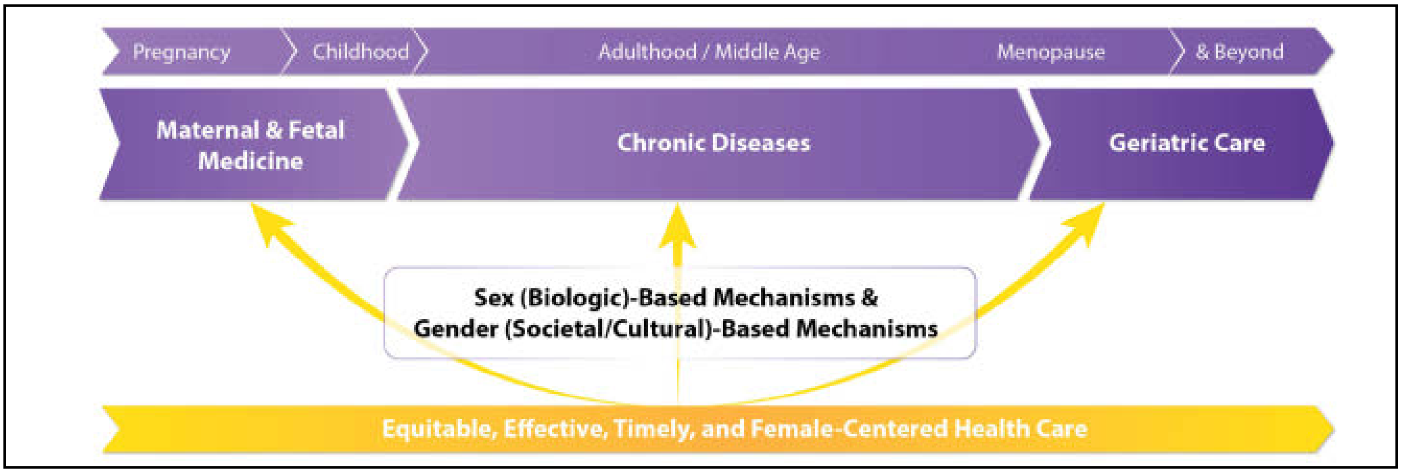

Pregnancy
Childhood
Adulthood / Middle Age
Menopause
& Beyond
Maternal & Fetal Medicine
Chronic Diseases
Geriatric Care
Sex (Biologic)-Based Mechanisms & Gender (Societal/Cultural)-Based Mechanisms
Equitable, Effective, Timely, and Female-Centered Health Care

**FIGURE 1. A multidisciplinary approach to women's health across their lifespan.**

- Today's health care environment is siloed and creates dramatic inefficiencies and discontinuity of care.
- By focusing on a woman's lifespan, unique cross-disciplinary teams of experts are brought together to research disease states and interconnecting factors impacting women throughout their lives.
- Connections forged across medical specialties enhance our understanding of how each stage of life contributes to women's future health and life expectancy.
- Societal, economic, and other disadvantages uniquely contribute to worsening disease states, diminished quality of life, and shortened life expectancy for women, especially for women of diverse race and ethnicity.
- Both sex (i.e., biologic) and gender (i.e., societal/cultural) factors uniquely contribute to varied female phenotypes and affect unique disease mechanisms and presentations, requiring female-centered care management strategies.

(Purple, white, and gold are the colors of the U.S. suffrage movement.)

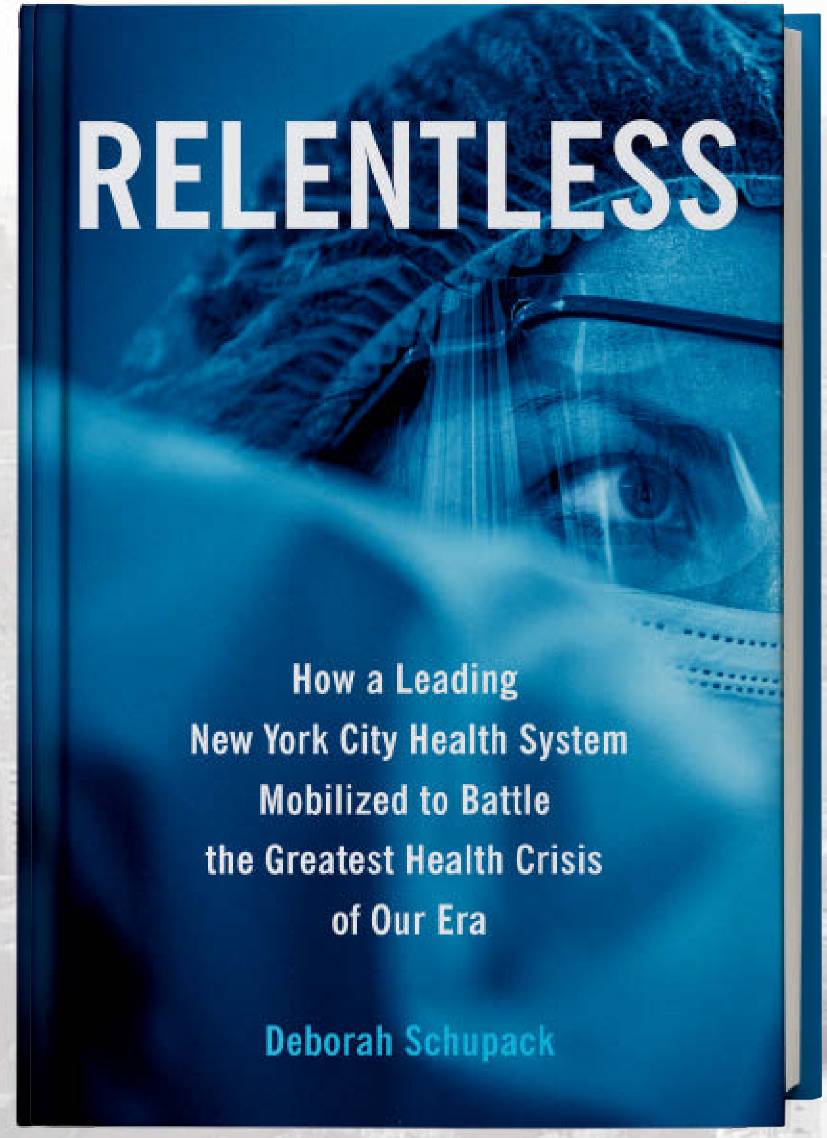
often based on the principle of "sameness" (*14*), where men and women have similar odds of prompt diagnosis and receipt of guideline-directed treatment. In reality, there is a prominent selection bias in referral of women, delayed presentation across varied diseases, suboptimal enrollment of women in clinical research, and underuse of a wide range of effective medical therapies and surgical procedures (*14*). Recent developments in artificial intelligence methods (*15*) are well suited to unearth the complexities of sex and gender science. Yet to be discovered is the way in which sex and gender features form complex patterns leading to varied disease manifestations that may be extracted by sequential predictive modeling using learning algorithms.

Targeting equitable care of women requires integrating gendered structural determinants of health with targeted sex-specific needs of women in order to define care coordination (*14*). Our most vulnerable racially and ethnically diverse women, lacking financial means and burdened with multigenerational care responsibilities coupled with societal discrimination, face a daunting task of seeking care in our health care system. People from sexual and gender minority populations also experience significant access and treatment barriers. Our health care system lacks impetus to change without transformative health policies, including public reporting of care for women of diverse race and ethnicity. For women's health scientists, innovation requires questioning everything and thinking creatively about diverse sex-based mechanisms, which include exploring novel strategies of care focused on early detection and averting the long-term adverse consequences leading to chronic diseases and premature mortality. The possibilities are abundant for scientists to realize major strides in improving the lives of women. It is vital that funding agencies fulfill their obligations of equitable enrollment and definable strategies of care to reduce the incidence of preventable illness and reduced life expectancy observed for too many women.

## References

1. T. Janevic *et al.*, *Health Aff. (Millwood)* **39**, 768–776 (2020).
2. T. Janevic *et al.*, *Matern. Child Health J.* **25**, 860–869 (2021).
3. J. Gofus *et al.*, *Scand. Cardiovasc. J.* **53**, 141–147 (2019).
4. E. Emmerson, M. J. Hardman, *Biogerontology* **13**, 3–20 (2012).
5. R. A. Lobo *et al.*, *Climacteric* **17**, 540–556 (2014).
6. B. Vogel *et al.*, *Lancet* **397**, 2385–2438 (2021).
7. M. Walker, M. Jacobson, M. Sobel, *CMAJ* **191**, E886–E893 (2019).
8. M. Merad *et al.*, "Building the Future of Immunology in Medicine," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 21–23.
9. S. M. Hayter, M. C. Cook, *Autoimmun. Rev.* **11**, 754–765 (2012).
10. H. Schlenz *et al.*, *Int. J. Obes. (Lond.)* **38**, S26–S31 (2014).
11. N. D. Wong, J. Pio, R. Valencia, G. Thakal, *Prev. Cardiol.* **4**, 109–114 (2001).
12. A. Khera *et al.*, *J. Am. Coll. Cardiol.* **46**, 464–469 (2005).
13. M. E. Ramos-Nino, *ISRN Oncol.* **2013**, 697521 (2013).
14. L. J. Shaw *et al.*, *J. Am. Coll. Cardiol.* **70**, 373–388 (2017).
15. T. J. Fuchs *et al.*, "The Advent of Artificial Intelligence in Health Care and its Practical Implementation," in *The Frontiers of Medical Research* (*Science*/AAAS, Washington, DC, 2021), pp. 4–5.

Icahn School
of Medicine at
**Mount**
**Sinai**


RELENTLESS
How a Leading
New York City Health System
Mobilized to Battle
the Greatest Health Crisis
of Our Era
Deborah Schupack

This book can be
pre-ordered now from
online booksellers and
will be in bookstores
early next year.

RELENTLESS is the story of how one of America's leading health care systems — the Mount Sinai Health System in New York City — mobilized to confront the defining health crisis of our time: the coronavirus pandemic.

This book weaves a narrative based on interviews with more than 100 people, including Mount Sinai's renowned research scientists who worked to decipher the virus and find ways to counter it.